FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

IPATIMUP
Instituto de Patologia e Imunologia Molecular da Universidade do Porto

MSc

A genetic approach to the relationship between taste perception and lifestyle in Africa

Marisa Sofia Sousa Oliveira

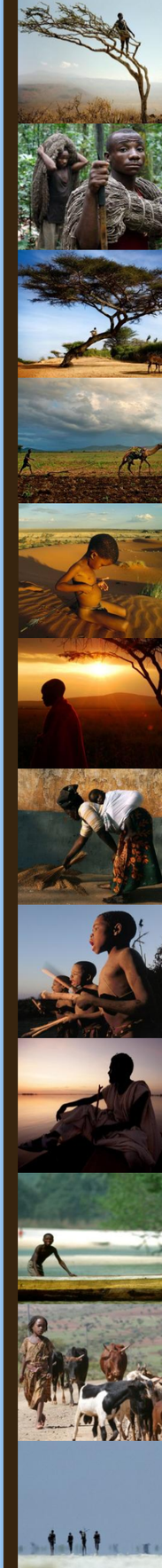
2012

FC

A genetic approach to the relationship between taste perception and lifestyle in Africa

Marisa Sofia Sousa Oliveira

2012



A genetic approach to the relationship between taste perception and lifestyle in Africa

Marisa Sofia Sousa Oliveira

Mestrado em Genética Forense

Departamento de Biologia

2012

Orientadora

Professora Doutora Maria João Prata Martins Ribeiro, Professora Associada c/ Agregação na Faculdade de Ciências da Universidade do Porto e Investigadora no Instituto de Patologia e Imunologia Molecular da Universidade do Porto.

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Dissertação de candidatura ao grau de Mestre em Genética Forense submetida à Faculdade de Ciências da Universidade do Porto.

O presente trabalho foi desenvolvido sob a orientação científica da Professora Doutora Maria João Prata Martins Ribeiro e foi inteiramente realizado no Instituto de Patologia e Imunologia Molecular da Universidade do Porto.

Dissertation for applying to a Master's Degree in Forensic Genetics submitted to the Faculty of Sciences of the University of Porto.

The present work was developed under the scientific supervision of Professor Maria João Prata Martins Ribeiro and was entirely done in the Institute of Molecular Pathology and Immunology of University of Porto.

Agradecimentos

Em primeiro lugar queria agradecer à minha orientadora, Professora Maria João Prata, por me ter dado a oportunidade de participar neste projecto, prontificando-se sempre a acompanhar-me e a ensinar-me. Agradeço o apoio e auxílio prestados no decorrer deste ano e a simpatia com que sempre me recebeu.

À Cristina e à Sofia, pela forma amigável como me acolheram desde o início e pela paciência e compreensão que tiveram durantes todas as fases, além da motivação que me transmitiram. Um agradecimento em particular à Cristina por toda a orientação e conselhos dados.

Ao Luis estou grata pela disponibilidade e contribuição que deu para a realização desta tese.

Ao Professor Amorim pela oportunidade de integrar o seu grupo e por tudo o que me ensinou ao longo dos últimos 2 anos.

Ao Grupo da Genética Populacional, pela forma agradável com que me recebeu. Além disso, a todos os outros que contribuem para o bom funcionamento e ambiente do IPATIMUP.

Aos colegas de mestrado por terem-se mostrado mais que simples colegas, principalmente as meninas Ana, Lídia e as redescobertas Filipa e Sofia, pela companhia e boa disposição diárias. Além disso, às duas aquisições vindas de Aveiro, Inês e Catarina, que vieram animar os nossos dias.

Aos amigos de licenciatura, que sem dúvida contribuíram para ter chegado aqui. Mas não posso deixar de referir em particular a Susana, Sara e Catarina, que se tornam aquele tipo de amigas que se quer manter durante a vida. E ainda, um agradecimento especial à Catarina, pois não só me acompanhou durante a licenciatura mas também no mestrado, e em ambos os casos ajudou-me (e aturou-me) como poucos. Obrigada Catarina!

Aos meus meninos, Ricardo e André, que me fazem rir em todas as ocasiões e conseguem tornar o pior dia num bem mais agradável…Obrigada por estarem presentes e serem assim e, sem dúvida, pela grande amizade que fomos construindo!

Às minhas meninas, Mónica e Tânia, pela amizade constante e por todos os momentos partilhados, bons ou maus, que nos fizeram crescer e sem dúvida contribuíram para o que hoje somos…Obrigada MarMoTa! (E ao Rui também!)

À Sofia, por tudo! Porque tens sido uma amiga excecional e sabes bem o valor que tens para mim…

Ao Prof. Zé Carlos, por ter sido muito mais que um professor mas um verdadeiro amigo e exemplo de vida. Por me ter dado importantes lições de Matemática na escola e no dia a dia e também por me ter ensinado a ser melhor e ter acreditado em mim. Por me ter feito “querer e crer” mais, não só como aluna mas como pessoa e pela fantástica oportunidade que foi conhecê-lo de perto e acompanhá-lo nos momentos difíceis. Nem todos tiveram essa oportunidade e eu não podia estar mais grata por pertencer a esse grupo. Obrigada Professor, até um dia!

À minha família e ao Jorge, por todo o apoio, não só neste percurso, mas em todos os momentos.

Aos meus avós, Eduardo e Margarida, por tudo o que fizeram por mim desde sempre, pela companhia, bons momentos e por me terem ensinado o verdadeiro significado da palavra “família”. Acredito que poucos têm uns avós como vocês, e eu estou orgulhosa de ser vossa neta!

À minha mãe…uma verdadeira lutadora e, acima de tudo, a minha melhor amiga. Obrigada por tudo e sei que, o que sou hoje, a ti o devo!

ABSTRACT

Diet is thought to have been a main factor determining the current worldwide patterns of human genetic diversity. Human evolution is associated to many dietary shifts, among which one of the most important happened around 10 000 YBP, when most human societies began to abandon hunting and gathering, replacing the subsistence strategy for agriculturalist and pastoralist lifestyles. The introduction of new food behaviours can imply changes in biological processes, and so diet might have acted as selective pressure leading to genetic alterations related not only to food metabolism but also to taste sensitivity given the role of genetics in certain food preferences. Concerning taste sensitivity, it is widely recognized that there are five basic taste qualities – salty, sour, sweet, bitter and *umami*. Despite being acknowledged that individual differences in taste sensibility are in part genetically determined, the knowledge about such relationship remains scarce.

The main goal of the present work was to characterize African populations with distinct modes of subsistence, using a selected battery of SNPs related to taste, in order to investigate whether patterns of diversity at those variations could represent genetic adaptations to lifestyle. Viewing that, three agrarian societies – from Angola, Mozambique and Equatorial Guinea – and one pastoralist – from Uganda – were genotyped with two multiplex reactions, designed and optimized to detect eleven SNPs allocated in four genes known to influence taste perception: *TAS1R1* and *TAS1R3*, associated with umami and sweet tastes; *TAS2R16* and *TAS2R38*, linked to bitter perception. In addition, samples from Portugal were genotyped to function as a control population from outside Africa.

For the screened variations no departures from Hardy-Weinberg equilibrium were detected in any of the population samples studied. Concerning *TAS2R16* gene, it was observed a clear differentiation between African and non-African populations, and within Africa, a remarkable parallelism was found between the allelic frequency distribution of the tested SNP at the gene and the Bantu expansion.

For most of the tested variations, AMOVA results indicated that geography was a main factor responsible for their worldwide patterns.

In this study, no clear association emerged between lifestyle and taste-related polymorphisms. Still, the scarcity of genetic data for African populations, does not allow to exclude that lifestyle has played a role in shaping diversity at the tested variations.

The worldwide patterns of genetic distribution of some of those variations revealed intriguing features suggesting that they might have been under selective forces, whose nature is yet unknown.

The present study has contributed to deepen the characterization of genetic diversity in Africa, opening new doors to achieve a better knowledge on the factors underlying the patterns of taste-related diversity in human populations.

Keywords: Diet, Lifestyle, Taste perception, Africa, Genetic Diversity.

RESUMO

Admite-se que a dieta tenha sido um importante fator modelador dos padrões atuais de diversidade genética humana. A evolução humana está associada a muitas mudanças na dieta, entre as quais se salienta a que ocorreu à volta de 10 000 anos, quando as sociedades humanas começaram a abandonar a caça e recoleção em benefício da agricultura e pastorícia como novos modos de subsistência. A introdução de novos comportamentos alimentares pode acarretar alterações nos processos biológicos, conduzindo a pressões seletivas, tanto a nível do metabolismo como na sensibilidade ao gosto. Quanto ao gosto, atualmente são reconhecidas cinco qualidades básicas – amargo, doce, salgado, ácido e *umami.* Apesar de se saber que as diferenças a nível individual na sensibilidade aos diferentes sabores são parcialmente reguladas pela genética, ainda pouco se sabe acerca dessa relação.

O trabalho apresentado surge com o principal objetivo de caracterizar populações africanas com distintos estilos de vida usando uma bateria de *SNPs* relacionados com o gosto. Neste sentido, três sociedades agrárias – de Angola, Moçambique e Guiné Equatorial, e uma pastoril – do Uganda, foram genotipadas com recurso a duas reações Multiplex especificamente desenhadas e otimizadas, contendo 11 polimorfismos implicados na perceção do gosto e distribuídos por quatro genes: *TAS1R1* e *TAS1R3*, envolvidos na perceção do *umami* e doce, e *TAS2R16* e *TAS2R38*, ambos associados à deteção do amargo. Além destas populações, foi caracterizada uma amostra de portugueses, que funcionou como um controlo de uma população fora do continente africano.

Quanto aos polimorfismos estudados, não foram encontrados desvios ao equilíbrio de Hardy-Weinberg em qualquer das cinco populações. Relativamente ao gene *TAS2R16*, observou-se uma clara distinção entre populações africanas e não africanas e em África encontrou-se um interessante paralelismo entre a distribuição de frequências do *SNP* testado nesse gene e a expansão Bantu.

Para a maioria das variações estudadas, os resultados da *AMOVA* revelaram que a geografia era um importante determinante dos respetivos padrões genéticos encontrados a nível mundial.

Neste trabalho, não se detetou qualquer associação clara entre estilo de vida e padrão de variação genética relacionada com o gosto. Contudo a escassez de dados sobre populações africanas com diferentes modos de subsistência, não permite ainda excluir que o estilo de vida tenha contribuído para configurar a diversidade genética das variações em análise.

Os padrões mundiais de distribuição de frequências quanto a algumas dessas variações, apresentam características peculiares, sugestivas de que possam ter estado sujeitas a efeitos de pressões seletivas, cuja natureza ainda se desconhece.

O presente estudo contribuiu para enriquecer o estado de caracterização genética em África, abrindo portas para se chegar a uma melhor compreensão dos fatores que influenciam os padrões de diversidade genética com impacto na sensibilidade gustativa.

Palavras-chave: Dieta, Modo de Subsistência, Sensibilidade ao Gosto, África, Diversidade Genética.

Table of Contents

Tables Index

Figures Index

Abbreviations

aa	amino acid
ALFRED	Allele FREquency Database
AMOVA	Analysis of Molecular Variance
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-like Alignment Tool
bp	base pair
ddNTP	dideoxynucleotide triphosphate
DNA	Deoxyribonucleic Acid
dNTP	dideoxynucleotide triphosphate
GMP	Guanosine 5'-monophosphate
GPCR	G protein-coupled receptor
He	Expected Heterozygosity
Ho	Observed Heterozygosity
IMP	Inosine 5'-monophosphate
LGM	Last Glacial Maximum
MDS	Multidimensional Scaling
µL	microlitre
mL	millilitre
Min	minute
MSG	monosodium glutamate
NCBI	National Center for Biotechnology Information
PCR	Polymerase Chain Reaction
PROP	6-n-propylthiouracil
PTC	phenylthiocarbamide
RFLP	Restriction Fragment Length Polymorphism
rpm	rotations *per* minute
SAP	Shrimp Alkaline Phosphatase
Sec	second
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
UCSC	University of California, Santa Cruz
YBP	Years before Present

I. INTRODUCTION

1. Population Genetics

Population Genetics is an essential discipline to a wide range of biology-related areas, from forensic and clinical genetics to evolutionary biology or anthropology.

Concerning human populations, the principal aim of Population Genetics is to obtain insights into the populations' history and understand how evolutionary and demographic events interacted leading to populations' changes and differentiation between each other. To achieve this goal, it is necessary to cross genetic information with data provided by historical, linguistics and archaeological records.

1.1. Selection and Genetic Adaptation

Analysing the diversity of human populations, there are indications that local adaptations have evolved improving human life under certain environmental conditions. Many of these adaptations have a genetic origin and take advantage on the standing variation created by mutation events.

Selection has an important role in shaping patterns of diversity. The survival of some genotypes can depend on the action of selective forces which, presumably, conduct to differences in the fitness of phenotypes. So, it is comprehensible that some individuals have advantage in relation to others, leading to preferential reproduction and gradual adjustment of the frequency of the more suitable genotypes in a population. The process of selection can lead to the fixation of advantageous mutations, to the elimination of deleterious alleles, but can also lead to the maintenance of multiple alleles in a *locus* at polymorphic levels. When a specific allele is associated to a favourable trait, its frequency tend to rise with generations, and concurrently other alleles decline in frequency more or less rapidly (directional selection). However, selection might also favour heterozygous individuals and, in this scenario, two or more alleles can persist in populations with polymorphic frequencies (Griffiths *et al.*, 2008).

It should not be forgotten, however, that the selective pressures determining the fitness of a certain genotype might vary with time and numerous environmental factors.

1.1.1. Selective Factors

Much speculation exists on the evolutionary impact of selective pressures in shaping current patterns of human genetic diversity. Up to now, evidence for genetic adaptations remains scarce, in part because it is difficult to prove that patterns of diversity at *locus* are non-neutral. Most of the times is not easy to distinguish whether the patterns are due to selection rather than demography (Balaresque *et al.*, 2007).

Broad patterns of global diversity are well explained by geography, based on the Out-of-Africa model of early human evolution (Balaresque *et al.*, 2007). The evident loss of genetic diversity with the increasing distance from the African continent is the outcome from the initial Out-of-Africa of *Homo sapiens* and sequential bottlenecks preceding expansions to other continents. It is well documented that Africa is the continent with greater diversity, supporting that was there that modern humans appeared.

In the recent years, the detection of signs of selection has been a major concern. Projects such as HapMap allowed a large scale access to human genome information. An extensive database of common single nucleotide polymorphisms (SNPs) was created, which facilitated studies aiming the identification of polymorphisms under selection (International HapMap Consortium, 2010). Many SNPs have already been analysed, related to distinct functions or characteristics, as for instance those implicated with cold tolerance, lactase persistence, skin pigmentation or resistance to malaria. Adaptations and consequent signals of selection are prompted by factors such as climate, diet and disease, which might vary from region to region (Balaresque *et al.*, 2007).

Among the presumed selective pressures, diet has recently owned special attention once the genome “stress” it imposes, is thought to be still important in nowadays societies.

The last major dietary modification in human evolution occurred in the Neolithic and continues to intrigue population geneticists (Ye and Gu, 2011). If shifts in food availability and diet composition likely created selective pressures on multiple biological processes, it remains unclear which potential genetic adaptations emerged since then. Recent studies performed to better understand the issue have identified in certain genes signals of being under selection, opening the way to more works concerning the influence of diet in the genetic diversity of contemporary human populations.

2. Diet

Diet is an important aspect of everyday life of human beings. It has changed significantly along hominids evolution, since the split between chimpanzees and early stages of hominids. Probably, major shifts were accompanied by genome adaptations to the new foods and dietary patterns appearing through times. Many anthropological studies have been devoted to elucidating this dietary history, in part because these shifts were likely associated with major anatomical and cultural changes, but also because this reconstruction is also crucial for understanding the evolutionary context of our modern diets and the diseases often associated with them (Luca *et al.*, 2010).

Concerning human past, three major dietary transitions occurred (figure 1). The first began with the separation among humans and chimpanzees (Chen and Li, 2001). The second arose with the emergence of the genus *Homo* and stone instruments, when meat or its derivatives become to be important food resources (Milton, 2003). The third and most recent shift happened nearby 10 000 Years Before Present (YBP), when humans began to domesticate animals and plants. The rise of agriculture and pastoralism marks the transition to Neolithic (Diamond, 2002).

At least concerning the Neolithic changes in dietary habits, there are indications that they were followed by a period of health problems and diseases connected with the non-adaptation of human genome. Some believe that diet-related mal-adaptations are still present in actual societies (Ye and Gu, 2011).

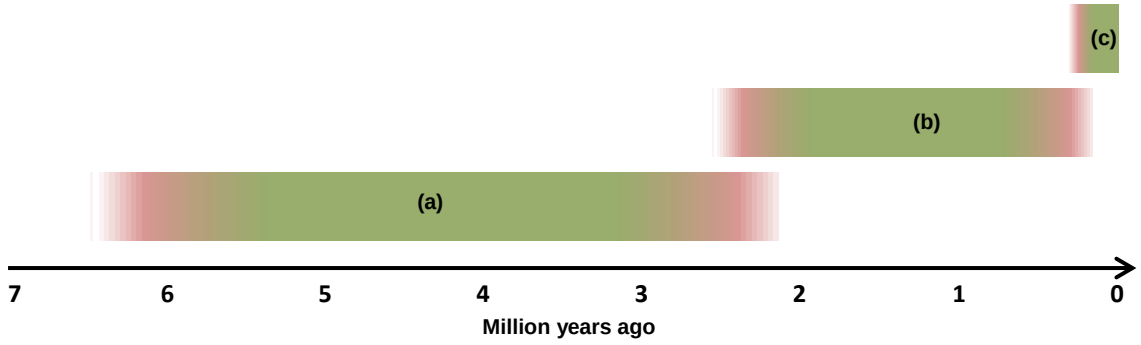


Fig. 1 – Schematic and simplified representation of the principal dietary shifts since the split between human and chimpanzee: **a)** plant-based diet, **b)** plant and meat-eating diet and **c)** domesticated and energy-rich diet. Adapted from Ye and Gu, 2011.

2.1. Neolithic Revolution

Nearly 200 000 years ago *Homo sapiens* appeared in Africa. The first modern humans were hunter-gatherers. Their feeding was dependent on many environmental variables, such as soil conditions, climate modifications and local fauna and flora. Hunter-gatherers' diet was mainly composed by local vegetable items, fruits and nuts provided by wild plants. In relation to the dietary meat component, sources as carrion and game were explored. In some groups, seeds and fruits were stored to support nutritional needs when other food sources were scarce. To maintain their populations' size, hunter-gatherers were nomad, moving to a new place when food supplies became limited, generally with season changes (Gupta, 2004).

All humans practiced this mode of subsistence till around 10 000 years ago, when the most important and recent dietary transition occurred associated to the emergence of agriculture (Ye and Gu, 2011).

Several factors might have triggered the appearance of agricultural and pastoralist lifestyles. A drastic change in climate conditions after the Last Glacial Maximum (LGM) likely accounted to this transformation, since animals, plants and humans were obligate, by drought circumstances resulting from LGM, to cohabitate closer benefitting from fertile lands and creating the necessary conditions to begin the process of domestication. By empirical artificial selection, controlled reproduction of species with specific traits, plants and animals with value and utility to humans were domesticated. As a result of the close contacts between human and animals, some diseases specific to animals passed to humans, which triggered the gain of defences against some animal diseases (Patin and Quintana-Murci, 2008).

The practice of agriculture and animal husbandry allowed the growth and settlement of populations at several focus in the world, since the ability to support larger groups in reduced areas increased comparing to hunting and gathering lifestyle (Diamond and Bellwood, 2003). Although the new mode of subsistence facilitated the storage and accumulation of food to be used in less favourable seasons, the change in dietary habits led to a decrease in diet diversity in comparison to that characteristic of hunter-gatherers', since only few species could be domesticated and cultivated. (Gupta, 2004; Luca *et al.*, 2010; Ye and Gu, 2011).

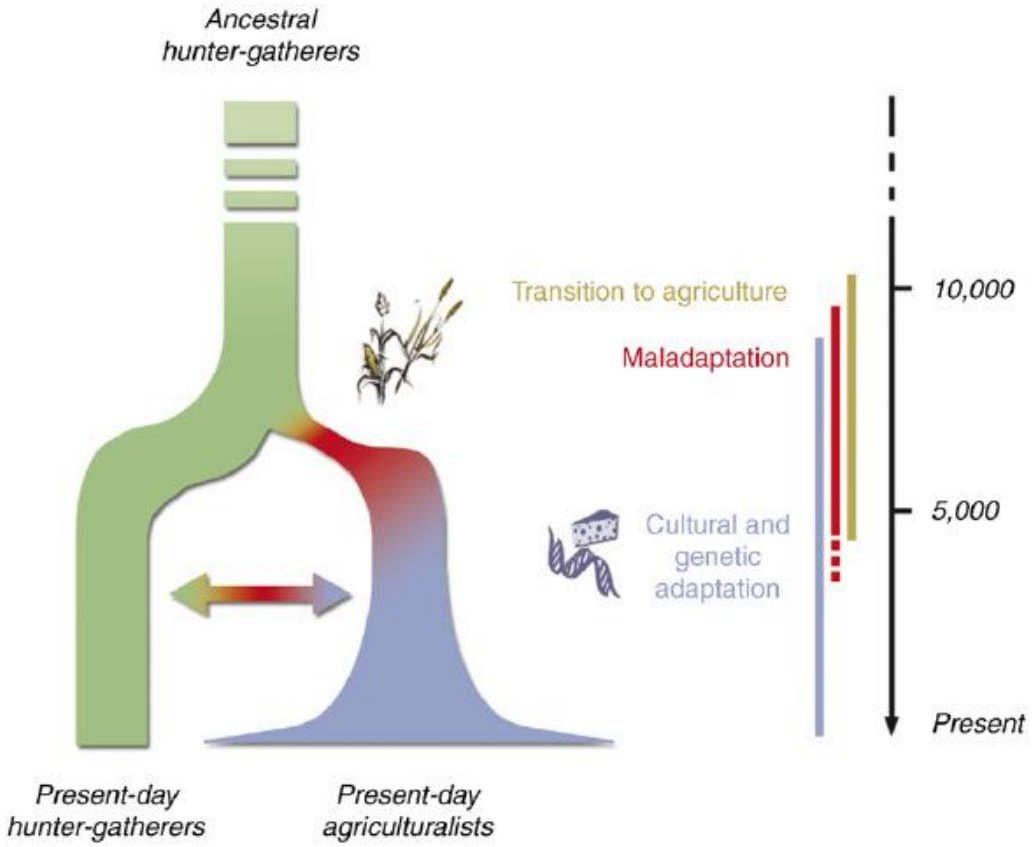


Fig. 2 – Representation of human lifestyles, diet and population size in the last 10 000 years. The first times of the transition to new dietary habits were characterized by the appearance of disorders caused by the non-adaptation of humans to them. Genetic and cultural alterations emerged, leading to agricultural population growth till present. Adapted from Patin and Quintana-Murci, 2008.

Despite the innovations brought by agriculture, like other major changes, it was also accompanied by complications. During the transition period (figure 2) occasional diseases as anaemia and metabolic disorders must have been common, once the organism and genome were not adapted to the new food resources (Jobling *et al.*, 2004; Balaresque *et al.*, 2007). The drastic dietary change, the new alimentary behaviours and sedentary lifestyle, led to the emergence of a series of associated diseases that progressively would turn into major health problems, like diabetes, cardiovascular syndromes, cancer and obesity (Feero *et al.*, 2010; Luca *et al.*, 2010). The nowadays burden of these diseases in most populations suggests that possible genome adjustments initiated in the last 10 000 years, do not fit the needs to deal with the current dietary practices and standards of living (Luca *et al.*, 2010; Ye and Gu, 2011).

Besides genetic adaptations, other answers to the Neolithic innovation were given with a repertoire of cultural adaptations (figure 2). Among them, food processing was very important once it enabled the consume of certain types of food decreasing their

toughness and neutralizing their toxins. Furthermore, the dissemination of techniques such as cooking and fermentation, allowed the demographic growth and the maintenance of larger groups (Patin and Quintana-Murci, 2008).

Agriculture did not have a single centre of origin. When it arose in Sub-Saharan Africa, it implied a major reshaping of the human landscape with the advent not only of farming communities but also of the pastoralist lifestyle. Besides, contrarily to what happened in most of the other continents, a few hunter and gatherer groups remained till present.

2.1.1. Diversity in Africa

Africa is one of the most interesting continents from the anthropological point of view. Besides having been the region where hominids and modern humans evolved, Africa has a long-lasting rich history of people movements. In the last centuries, the most impressive was the massive African slave trade initiated with the Age of Discoveries, during which also occurred some influx, although minor, of Europeans into Africa.

The continent possesses an extraordinary diversity: climatic, geographical (from deserts and mountain ranges to savannah and tropical rainforest) and even of infectious agents (some of them almost eradicated elsewhere). Furthermore, the highest levels of genetic diversity are present in African populations. Many distinct groups with a wide range of lifestyles can be found in Africa, hunter-gatherers, agriculturalists and pastoralists, notwithstanding the massive demographic impact of the farmer expansions that begun in Neolithic around 4000 YBP (Reed and Tishkoff, 2006; Campbell and Tishkoff, 2008).

It is also very high the linguistic diversity. Languages spoken in Africa belong to different families (figures 3 and 4), three of which can be related to distinct lifestyles. Within the Niger-Kordofanian branch, are the Bantu languages, which are associated with the spread of agriculture. The Nilo-Saharan family contains languages spoken by pastoralists' societies. The Khoisan branch is typical from hunter-gatherer groups; it is thought to be the most ancient linguistic family in Africa, which was dramatically abandoned due to the expansion of the Bantu languages during Neolithic (Tishkoff *et al.*, 2009; Campbell and Tishkoff, 2010).

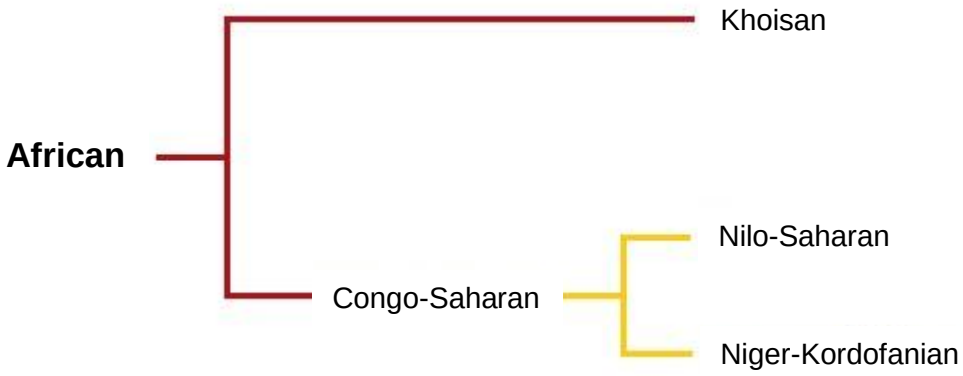


Fig. 3 – Principal linguistic families of Sub-Saharan Africa simplified. Adapted from Cavalli-Sforza, 1997.

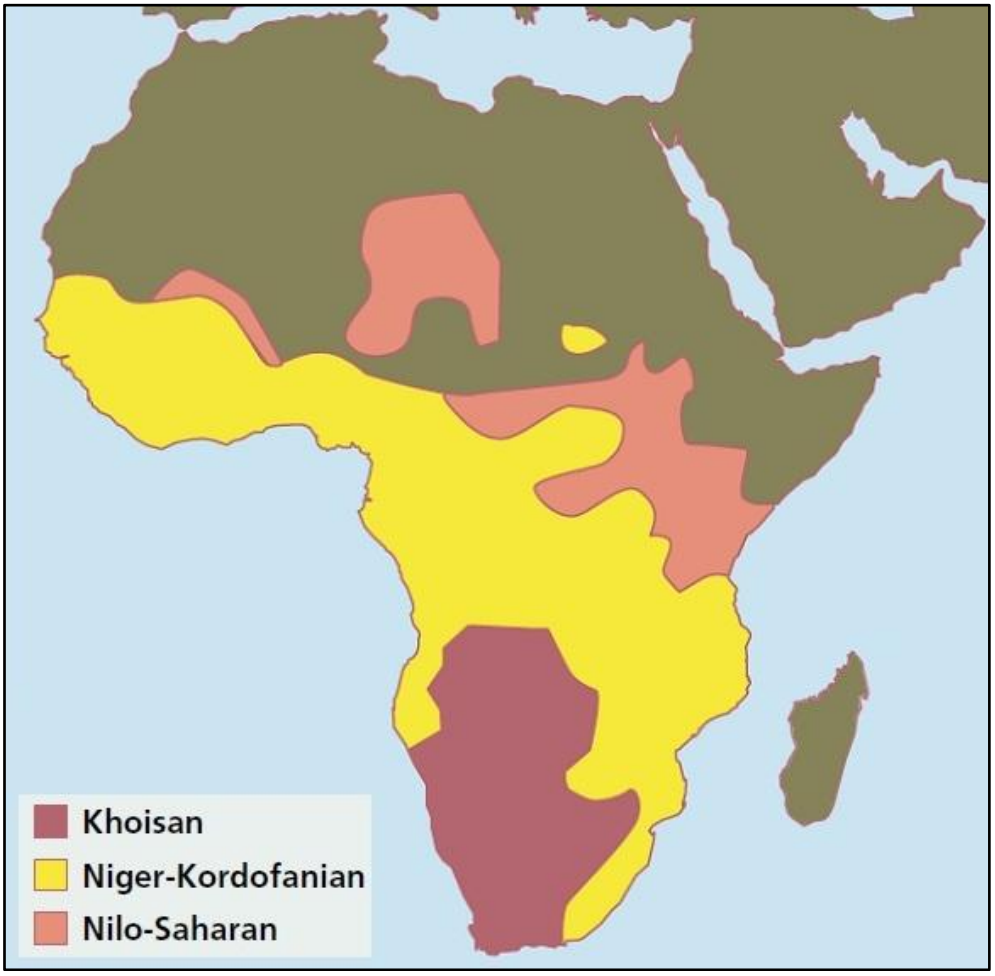


Fig. 4 – Principal languages families in Africa. The Niger-Kordofanian branch includes all the Bantu languages, which are primarily linked with agriculturalist lifestyle; the pastoral dispersion is associated with Nilo-Saharan languages and the remaining practitioners of hunting and gathering speak Khoisan-related languages. The area occupied by the language families is approximately the same filled by the related lifestyle. Adapted from Cavalli-Sforza and Feldman, 2003.

2.1.1.1. Agriculture and its expansion

Neolithic in Sub-Saharan Africa started around 5000 to 4000 YBP probably due to a severe change in climate from humid to drier, resulting in the introduction of new crops adequate to the changed conditions. Soon after, begun the Bantu expansions, a long series of migrations that would represent a remarkable process of farming/language co-dispersal (Diamond and Bellwood, 2003; Campbell and Tishkoff, 2010). The centre of origin of the diffusion of African farmers is widely thought to be an area around eastern Nigeria and western region of Cameroon (Diamond and Bellwood, 2003; Tishkoff *et al.*, 2009; Campbell and Tishkoff, 2010; Pakendorf *et al.*, 2011; Gignoux *et al.*, 2011). From there, two distinct dispersal routes arose: one throughout western Africa and another towards central and eastern regions of the continent, both in direction towards south. As a result, within Bantu people, western subgroups tend to be somehow more diverse than eastern ones (Pakendorf *et al.*, 2011). Furthermore, while most languages from Western Bantus is classified as “Forest” languages, those from Eastern subgroups are most often “Savannah” languages (Holden, 2006; Campbell and Tishkoff, 2010; Pakendorf *et al.*, 2011).

The agricultural expansion was highly correlated with the spread of Bantu languages through Sub-Saharan African. As a consequence of the massive co-dispersion, pre-existing hunter-gatherers and their culture were rapidly replaced and/or assimilated by the incomers Bantu-speakers. Evidence exists that intermarriage between Bantu and Khoisan people was a sex-biased process, having involved preferentially mattings between hunter-gatherer women and agriculturalist men (Diamond and Bellwood, 2003).

Tracing the farming expansion routes has been a difficult endeavour, due to the complexity of the movements, which is further complicated by the fact of being quite common situations of multilingualism. Nowadays, however, the geographical distribution of Bantu languages and speakers vastly dominates the sub-Saharan Africa. (Berniell-Lee *et al.*, 2009; Campbell and Tishkoff, 2010; Pakendorf *et al.*, 2011).

2.1.1.2. Pastoral migrations

Another lifestyle that emerged in Africa during the Neolithic was pastoralism. This mode of subsistence essentially relies on domestic herd products such as milk, blood and meat. Naturally, the main activities in pastoralist groups are related to animal husbandry. This lifestyle is deeply correlated with Nilo-Saharan languages spread among eastern and central Africa.

A major original region from where herders become to migrate is thought to have been in the south of Sudan (Campbell and Tishkoff, 2010). Movements of small groups of herders to southern regions might have been coerced by the enlargement of arid areas surrounding the Sahel desert (Marshall, 1990; Campbell and Tishkoff, 2010).

In Africa, pastoralists often complement herd products with others obtained through hunting, fishing, cultivation and foraging, which affords a richer diet. This form of pastoralism is normally named generalist, having appeared at least 4000 YBP. Still, in East Africa other varieties of pastoralists are present that are more specialized, living almost exclusively from herder labour and excluding all meat sources from wild nature (Campbell and Tishkoff, 2010). This form of pastoralism is considered to be more recent, arising around 3000 YBP, probably to cope with the climatic conditions in that time (Campbell and Tishkoff, 2010).

Climatic changes seem had played a major role in driving the Nilo-Saharan expansion to surrounding regions, initially to north Kenya and Lake Chad. Still, a much smaller group expanded in direction to eastern Sahara (Campbell and Tishkoff, 2010). Around 3000 YBP in the late Holocene occurred another climate shift into a drier weather that resulted in a reduction of humid areas, such as lakes and rivers, facilitating the acclimation of cattle to these new territories. Then, around 3000 YBP, a distinct group from the east side of Sudan initiated their expansions to south, through routes essentially leading them to Uganda, Kenya and Tanzania regions. Nowadays they are denominated Nilotic pastoralists (Marshall, 1990; Bower, 1991; Campbell and Tishkoff, 2010).

2.2. Dietary Adaptations resulting from Neolithic innovations

Increased population densities, reduced dietary diversity, sedentary lifestyle and exposure to animal pathogens, together represented a major set of challenges appearing with the Neolithic, some of which may still have an impact today. A common view is that the arrival of agriculture signalled the start of an era of dietary maladaptation and to face with that humans begun answering with a series of genetic adaptations (Balaresque *et al.*, 2007).

To elucidate which genetic adaptations might have evolved in response to dietary changes and the adoptions of dietary specializations, numerous genes have been investigated, especially those involved in food metabolism (Nielsen, 2005).

Up to now, one of the most well evidenced cases of genetic metabolism adaptation refers to the human tolerance to lactase in later stages of life, regarding which groups of African pastoralists are clearly differentiated from non-pastoralist groups (Hollox *et al.*, 2001; Powell *et al.*, 2003). Copy number variation of the *AMY* gene was also associated with starch intake and lifestyle in human populations, accounting for differences between agriculturalists and hunter-gatherers (Perry *et al.*, 2007). These findings support that some local adaptations exist which are correlated with dietary behaviours and main modes of subsistence adopted (Campbell and Tishkoff, 2010; Ye and Gu, 2011).

However, feeding is a multisensory experience. The processing of food in the mouth leads to the release of molecules that stimulate, among others, the sense of taste (Luca *et al.*, 2010). Taste is an important factor of food selection. Besides determining individual differences in food preferences, taste sensibility is very important to avoid the ingestion of substances like poisons or simply spoiled food (Bachmanov and Beauchamp, 2007). An additional function of taste is the detection of energetic food resources as a guarantee of individuals' continuity and their reproduction (Drewnowski and Rock, 1995; Kim *et al.*, 2004; Tepper, 2008).

As so, a few studies appeared recently sought to understand whether taste in humans might also represent a genetic adaptation to dietary changes (Luca *et al.*, 2010).

3. Taste Perception

In general, the term “taste” has been applied to all sensations that come from oral cavity. However, its biological meaning only comprises the sensations originated from the stimulation of taste receptor cells by the chemical particles present in food consumed. The taste system comprises several complex and specialised structures that allow the identification of different types of taste (Bachmanov and Beauchamp, 2007; Luca *et al.*, 2010).

Nowadays are widely recognized five basic tastes: bitter, *umami*, sweet, salty and sour (Drewnowski, 2002), although new qualities of taste have been proposed in recent years such the ability to taste water and fat (Laugerette *et al.*, 2005; Laugerette *et al.*, 2007). Each of the 5 basic tastes is associated to specific taste receptors that transmit the signal to the brain in order to obtain the corresponding taste perception (de Krom *et al.*, 2009).

Despite the number of studies performed in the last years to better understand the mechanism of taste perception, it is still poorly known how genetic variation influences taste perception (Wooding, 2005).

3.1. Bitter Taste

The bitter perception has presumably evolved to avoid the ingestion of plant toxins, provoking an unpleasant sensation in mouth which induces their rejection. Bitter compounds have high molecular diversity and are found in a wide range of plants such as cruciferous vegetables, spinach and endives. Furthermore, its presence is detected in other non-natural types of food, such for example certain cheeses, products with soy, beer and coffee (Keller *et al.*, 2002; Garcia-Bailo *et al.*, 2009).

The sensibility to bitter has been intensively study since the 1930s, after an interesting observation made by Arthur L. Fox, during his work with a compound named phenylthiocarbamide (PTC) and its related 6-n-propylthiouracil (PROP). Testing PTC sensitivity in a large population sample, he discovered two principal phenotypic classes of perception: one contained people who were sensible to low concentrations of PTC compound, denominated tasters, and the other those who only tasted PTC at high concentrations, the non-tasters.

Although PTC has being the more known and studied compound eliciting bitter taste, it is not naturally present in food, but it triggers a similar response to the one

obtained with isothiocyanates, isoflavones and polyphenols, just citing few examples (Wooding, 2006; Garcia-Bailo *et al.*, 2009; Tepper *et al.*, 2009; Feeney *et al.*, 2011).

The bitter receptors are encoded by a large family of seven-transmembrane G protein-coupled receptor (GPCR) genes named TAS2R. The gene family is composed by at least 38 functional genes and 5 pseudogenes which are organized in gene clusters located in three chromosomes, 5p, 7p and 12p, and it has evolved through gene duplication during mammalian evolution (Bufe *et al.*, 2005; Drayna, 2005; Tepper *et al.*, 2009; Ye and Gu, 2011).

TAS2R38 is the more known gene of this family because it is responsible for PTC sensitivity. Its length is around 1143 base pairs (bp), codes for 333 amino acids (aa) and its location is on chromosome 7 (http://www.ensembl.org). The PTC genotype is commonly accessed by studying three functional single nucleotide polymorphisms (SNPs), which are present in table 1. They define different haplotypes, among which are PAV and AVI haplotypes that are strongly associated to tasters and non-tasters phenotypes, respectively.

Table 1 – Single Nucleotide Polymorphisms linked to PTC perception in *TAS2R38*.

Reference SNP (rs)	Position in sequence (bp)	Possible Alleles (bp)	Position in sequence (aa)	Possible Alleles (aa)
rs713598	145	C G	49	Proline Alanine
rs1726866	785	C T	262	Alanine Valine
rs10246939	886	G A	296	Valine Isoleucine

Another gene from the family that is calling attention is *TAS2R16.* It is also located at chromosome 7, spanning 996 bp that encoding 291 aa (http://www.ensembl.org). This gene determines the sensitivity to bitter β-glucopyranosides, which are present in a wide range of plants and are characterised by high toxic cyanogenic activity. The receptor encoded by *TAS2R16* also mediates the signaling response to salicin and amygdalin. Moreover, the gene has been studied in the context of the behaviour of alcohol intake (Soranzo *et al.*, 2005; Garcia-Bailo *et al.*, 2009; Ye and Gu, 2011).

So far, only one SNP at *TAS2R16* (rs846664) – G516T – has been associated to increased β-glucopyranosides sensibility (Bufe *et al.*, 2002).

Both genes show signs of being under selection (Wooding, 2011). In the case of *TAS2R38*, it has been found to be under balancing selection due the excess of

intermediate frequency variants. Concerning *TAS2R16*, based on the very high frequency of some derived alleles involving nonsynonymous nucleotide changes, it was also proposed that they were putative targets of selection (Ye and Gu, 2011).

Nevertheless, the reasons underlying the detected selection signals remain unclear.

3.2. *Umami* Taste

The *umami* taste was the last to be included in the list of taste qualities and it is usually described as a savoury flavour. It was firstly identified by Kikunae Ikeda in 1908 (Lindemann *et al.*, 2002; Kurihara, 2009), hence the origin of the Japanese word *umami* that means good taste. It is used to describe the taste of the amino acid L-glutamate, usually present in food as monosodium glutamate (MSG), together with L-aspartate. Inosine 5'-monophosphate (IMP) and guanosine 5'-monophosphate (GMP) are responsible for increased MSG and L-aspartate *umami* gustation. The *umami* taste is characteristic of meat, fish, milk, cheese, mushrooms, potatoes, tomatoes, soy and seafood, among others (Kurihara and Kashiwayanagi, 2000; Chaudhari *et al.*, 2009; Garcia-Bailo *et al.*, 2009).

The *umami* receptors are very complex, but also implicate G protein-coupled receptor genes, among which are *TAS1R1* and *TAS1R3,* which harbour several SNPs known to be correlated with *umami* sensibility. These genes, as well as *TAS1R2,* belong to the TAS1R gene family that is located in a single cluster on chromosome 1 (Li *et al.*, 2002; Zhao *et al.*, 2003). To enable the perception of *umami* taste, the proteins encoded by genes *TAS1R1* and *TAS1R3* that are predominantly expressed in taste buds need to form heteromeric taste receptors. Besides the referred *umami* receptor, two glutamate-selective GPCRs, mGluR1 and mGluR4, have also been proposed as candidate taste receptors for *umami* (Nelson *et al.*, 2002; Chaudhari *et al.*, 2009; Garcia-Bailo *et al.*, 2009; Shigemura *et al.*, 2009a; Shigemura *et al.*, 2009b).

Till this point, several SNPs were identified in both genes shown to affect *umami* sensitivity, having *TAS1R1* more polymorphisms than *TAS1R3*, which can be related with *TAS1R3* double function in the perception of *umami* and sweet tastes (Kim *et al.*, 2006; Garcia-Bailo *et al.*, 2009; Chen *et al.*, 2009). Despite that, the understanding about how genetic variations influence the sensibility to *umami* and food choice remains limited.

3.3. Sweet Taste

Food containing sweet substances usually provides a pleasant sensation in humans, possibly reflecting evolutionary pressures to select foods high in energy (Hladik *et al.*, 2002). Most often the sweet perception in humans is elicited by natural sugars, like glucose, fructose, sucrose, and sugar alcohols such as sorbitol. Several other natural compounds, structurally unrelated to carbohydrates, also taste sweet. Most commonly, certain amino acids, such as glycine, taste sweet to humans. Furthermore, artificial sweeteners have been developed activating similar responses to those obtained with natural sugars (Bachmanov *et al.*, 2002; Breslin and Spector, 2008; Boughter and Bachmanov, 2007; Garcia-Bailo *et al.*, 2009; Bachmanov *et al.*, 2011).

As *umami* taste, sweet is also dependent on proteins encoded by genes of TAS1R family and GPCRs. In this case, the genes involved are *TAS1R2* and *TAS1R3*, whose proteins form a heteromer enabling sweet identification in the taste buds of oral cavity. Comparative levels of gene variability, reveal that *TAS1R2* is notably more polymorphic than *TAS1R3* or *TAS1R1*, suggesting that variation at *TAS1R2* generates the ability to perceive a wide variety of structurally different sweet substances (Garcia-Bailo *et al.*, 2009). The *TAS1R3* gene has demonstrated strong association with differences in sweet taste perception of saccharin (Nelson *et al.*, 2002; Zhao *et al.*, 2003; Kim *et al.*, 2006).

Despite the discovery of these genes as coding for an heteromer responsible for sweet taste detection, the number of sweet taste receptors that exist is still unresolved (Garcia-Bailo *et al.*, 2009; Bachmanov *et al.*, 2011).

3.4. Salty Taste

The salty sense can be stimulated by several ions like Li^+, K^+ and NH_4^+, but the most effective stimulus of this sensation is NaCl (DeSimone and Lyall, 2006; Roper, 2007). NaCl is required to preserve the electrolyte equilibrium, regulate blood pressure and volume. Furthermore, NaCl provides ions that have crucial roles in many physiological mechanisms (Kim *et al.*, 2004; Chandrashekar *et al.*, 2006).

Salt recognition is thought to occur in sodium channels from the epithelium, although in humans other mechanisms were proposed to have a role in its perception. Even so, the molecular mechanism responsible for salt taste perception in humans

remains largely unclear. Furthermore, while genetic contributions to salt preference and consumption have received long-standing attention in humans, up to now, no genetic variation could be firmly connected with salty taste perception (Wise *et al.*, 2007; Garcia-Bailo *et al.*, 2009).

Results from a recent study indicate that environment plays a larger role than genetics in determining individual differences in recognition thresholds for saltiness (Wise *et al.*, 2007). So, it still needs to be uncovered how genetic variation may affect salt sensitivity in humans.

Given the importance of salty compounds to human life, it has been admitted that evolutionary forces had acted, shaping the variation in salt sensitivity observed today (Wise *et al.*, 2007; Garcia-Bailo *et al.*, 2009).

3.5. Sour Taste

It is though that sour sensibility may help prevent the consumption of spoiled foods or serve as an indicator of fruit ripeness (Garcia-Bailo *et al.*, 2009). The main sources of sour taste are organic acids such as lactic, citric and acetic, which are present in a great amount of vegetables and fruits, wine and some animal products.

Sour sense is activated when acid compounds after reaching taste buds cause depolarization-induced Ca^{2+} entry into taste receptor cells. However, the molecular machinery accounting for sour taste perception still remains poorly understood.

In the past few years, two ion channels, PKD2L1 and PKD1L3 from the polycystic kidney disease-like subfamily of transient receptor potential, have gathered strong evidence as putative sour taste receptors. Studies performed in mouse revealed that the proteins encoded by PKD2L1 and PKD1L3 functioned as heteromer in a different subset of taste cells than those for sweet, bitter, and *umami* (LopezJimenez *et al.*, 2006). Although this finding highlight the role of PKD2L1 and PKD1L3 in sour taste perception, the potential relationship between polymorphisms in these genes, sour taste perception, and subsequent food choices remains to be explored (Garcia-Bailo *et al.,* 2009).

Much is unknown about global patterns related to variation in sour taste perception, as well on how such variation is genetically determined. Nonetheless, two recent studies based on the analysis of twins, clearly demonstrated that genetic factors played a larger role than environment to the individual differences in sour perception

(Wise *et al.*, 2007; Tornwall *et al.*, 2012). So, many work must still be done to identify which genetic factors are those (Garcia-Bailo *et al.*, 2009).

4. Forensic Genetics Approach

Forensic Genetics plays an important role in nowadays society. DNA analysis has been used as an important tool to solve from simple cases of paternity to violent crimes such as rapes and murders. Around the world emerged databases at local level containing the information on a number of genetic markers located in non-coding regions in the genome to not enable the access to individual traits like genetic diseases, skin and eyes colour (Butler, 2005). Despite these concerns in not exposing an individual's phenotype, recently forensic geneticists gained particular interest in genes associated with physical traits. This new upcoming forensic discipline was named DNA phenotyping since it is aimed at helping to identify criminal suspects based on traits such as skin, hair and eye colour, or geographic ancestry, by analysing target DNA polymorphisms known to influence those characteristics. DNA phenotyping seeks to predict a criminal's appearance, which can be used to narrow the search for suspects, but using genetic information, assuming that this kind of information can be much more credible than that based on witness description.

Walsh *et al.* (2011a and 2011b) published a multiplex, denominated IrisPlex, designed to assess the iris colour. Through minisequencing technique, it was possible to determine the eye colour of an individual, with a high degree of confidence – more than 90% in the case of blue or brown eyes, based on a set of six single nucleotide polymorphisms. Furthermore, worldwide frequencies were accurately estimated, providing the first data to start constructing a new database with genetic information exclusively associated with phenotypic characteristics relevant in the forensic context. Other studies involving additional traits are still undergoing. In the case of hair colour, a preliminary work was also published in 2011 (Branicki *et al.*, 2011), reporting on 13 genetic markers distributed over 11 genes that enable the identification of hair colour with around 90% of accuracy for red, black and blonde colours and exceeding 80% for brown hair.

Despite the development and investment in this area during the last few years, little has been reached in terms of applicability in the forensic casework. The major obstacles are the new ethical issues raised and the current legislation of almost every country. In Europe, only Netherlands approved a law that regulates the use of DNA coding regions to access individual's phenotype to forensic purposes (Enserink, 2011).

If all precautions are taken to not use data incorrectly, the prospects of DNA phenotyping in forensic casuistic might be high. The implementation of standard laws to uniform its applicability will represent a priority step in the future. But in the

meanwhile, it is necessary to continue exploring the relationship between genotypes and phenotypes to convert a subject's information into a higher level of accuracy. Validating tests need to be rigorously performed before using genetic information to predict a phenotypic trait of an individual. Eventually, extensive databases will be constructed containing frequencies of genes influencing several phenotypic characteristics.

In this scenario, the study of taste perception related SNPs can represent a complementary tool in the field. Once being linked to phenotypes, the information afforded by taste-related genetic variations can contribute to increase the predictive knowledge on the phenotypic profile, or even behaviour, of an individual.

II. AIMS

In order to investigate the relationship between variability in taste perception and lifestyle, African populations with distinct modes of subsistence will be studied.

The interest in African populations relies in several reasons, among which is the fact that Africa has revealed to have the highest levels of diversity at worldwide scale (Campbell and Tishkoff, 2008). Despite that, the knowledge about genetic diversity in this continent remains scarce. Besides, the three main modes of subsistence – hunting-gathering, pastoralism and agriculture – are still represented in Africa, with fewer modifications than in other regions of the world.

After a brief review for each taste and their candidate genes, three qualities of taste were selected to be our focus: bitter perception, due to its importance to prevent toxin ingestion, and sensitivity to sweet and *umami*, both implied in detection of energetic sources (Feeney *et al.*, 2011).

To obtain insights on the main question under study, the present work aims at reaching the following objectives:

1. Design a PCR-Multiplex reaction selecting a battery of relevant SNPs, known to be implied in taste sensibility;
2. Characterize three African agrarian societies – Angola, Mozambique and Equatorial Guinea – and one pastoralist from Karamoja region (Uganda) for the previous selected SNPs;
3. Genotype an European control group, which will be represented by a sample of Portuguese, with the same battery of SNPs;
4. Explore the correspondent patterns of diversity and their relation with the different lifestyles of the studied populations;
5. Identify some of the factors that have influenced the detected patterns of genetic diversity.

In addition, the utility of these polymorphisms in the forensic genetics field, such as DNA phenotyping, will be evaluated, since they can be a useful complement to already used polymorphisms linked to phenotypic traits, increasing the discrimination power of the entire set.

III. MATERIAL & METHODS

1. Samples and DNA extraction

In order to perform the present work, samples from four African and one European country (figure 1) have been analysed. To characterize the agrarian societies we have used 27 samples of Bantu speakers from Angola, mainly from Cabinda; 34 from Mozambique; and 76 samples from Equatorial Guinea. To represent the pastoralist lifestyle, 49 male individuals, belonging to an ethnic group living in the north-east region of Uganda (Karamojong people) and who speaks an Eastern Nilotic language, have been typed. A control group comprising 49 samples from Portugal has been used as a reference from outside Africa. Once data about the dietary habits of the selected African populations were not collected during the sampling, their characterization was performed attending to linguistic family, region and other information obtain in literature.

The samples have been extracted previously for other works, through three different methods – Chelex®-100™ (Biorad) described by Lareu *et al.* (1994), phenol-chloroform (Maniatis *et al.*, 1989) and commercial Generation® Capture Card kit (Gentra Systems Inc, Minneapolis, USA) – following the standard protocols.

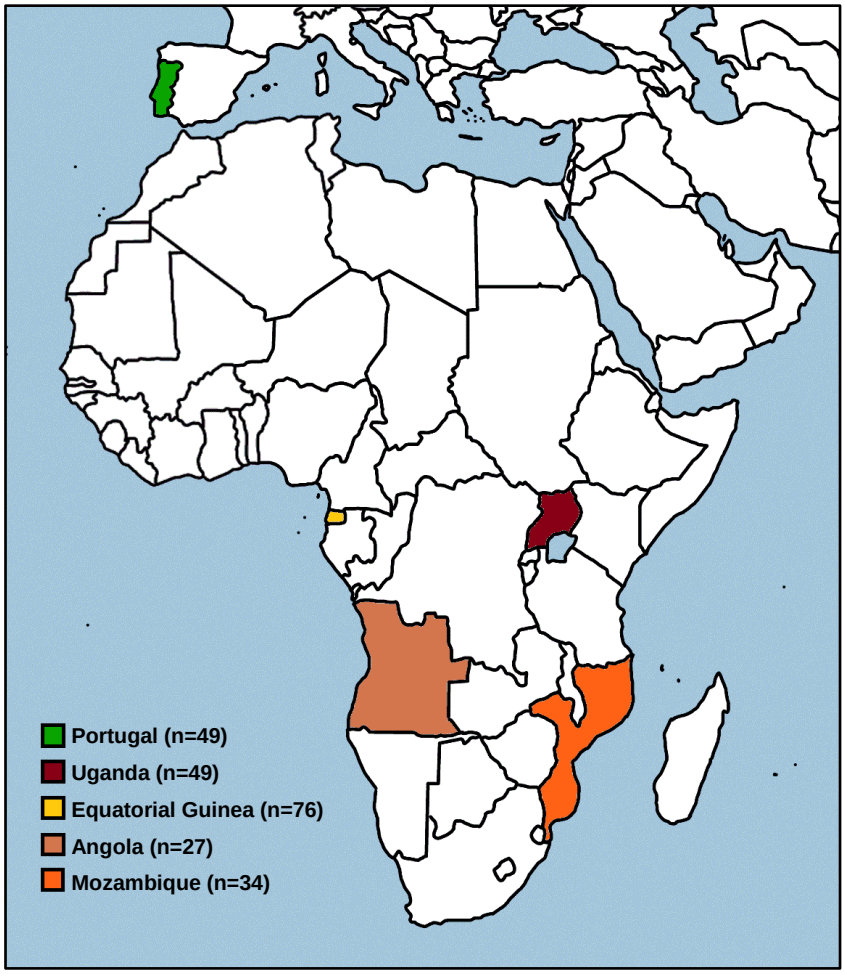


Fig. 5 – Representation of African continent and part of Eurasia with the sampled countries labeled.

2. Amplification Multiplexes Design

2.1. Genetic Markers and Target polymorphisms selection

One of the most important steps in this project was to select carefully the markers to be analysed. All the markers selected were single nucleotide polymorphisms, also known as SNPs.

To access the genomic information about taste singularities in pastoralist and agriculturalist lifestyles, a battery of eleven SNPs distributed across four taste-related genes have been studied. The genes selected were associated with bitter, *umami* and sweet tastes; this assignment was performed considering the relevance of polymorphisms described in the literature and its effects in taste perception.

In relation to *umami* taste, 2 SNPs from *TAS1R1* gene (rs41278020 and rs34160967) and 3 from *TAS1R3* (rs76755863, rs111615792 and rs307377) were selected.

As for the sweet taste, two SNPs from gene promoter of *TAS1R3* were chosen – rs307355 and rs35744813.

Concerning the bitter perception, the SNP rs846664 from *TAS2R16* gene and 3 SNPs from *TAS2R38* gene – rs713598, rs1726866 and rs10246939 – were included in the present work.

2.2. PCR Multiplex amplification

Viewing maximum efficiency in laboratory routine, nine pairs of primers have been designed to amplify in multiplex the specific DNA regions containing the target SNPs, which allows the amplification of a large number of fragments at the same time.

The gene sequences were retrieved from Ensemble Genome Browser (Ensembl 64 – Sep 2011) (http://www.ensembl.org) and used in Primer3 v.0.4.0 software (www://frodo.wi.mit.edu/) (Rozen and Skaletsky, 2000) to design the amplification primers avoiding annealing with polymorphic regions. To determine the specificity of the primers retrieved by Primer3, a double test was performed to each pair in Primer-BLAST, provided by NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), and UCSC *In-Silico* PCR (http://genome.csdb.cn/cgi-bin/hgPcr), choosing only the human genome. If any of the oligonucleotides anneals in other region(s) of the genome, it could imply a less efficient reaction or a non-specific amplification. Attending this, all the primers were tested individually in BLAST (NCBI) (http://blast.ncbi.nlm.nih.gov/) and BLAT (UCSC) (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start).

The last step of multiplex design consisted in using the AutoDimer software version 1.0 and OligoCalc (http://www.basic.northwestern.edu/biotools/OligoCalc.html) to check if the designed amplification primers could form primers dimers and/or hairpins, which could contribute to decrease the reaction efficiency. In table 2 are present all the amplification primers and SNPs of interest.

Table 2 – Amplification primers and their characteristics.

Gene	SNP	Allele	Forward Primer (5'->3')	Reverse Primer (5'->3')	bp Forward	bp Reverse	Product Size
TAS1R1	rs41278020	C329T	TCAATGAGCATGGCTACCAC	CACCGTAGGGGAATAGTGGA	20	20	218
TAS1R1	rs34160967	G1114A	CCTGAAGGCGTTTGAAGAAG	GGCAGAACTCATGGAGAAGG	20	20	160
TAS1R3	rs76755863	G13A	CCTGTTGGAAGTTGCCTCTG	ACGTAGTCCCCCTTCATCCT	20	20	129
TAS1R3	rs111615792	G740A	GGGCCTGAGCATCTTCTCG	ACCACCTGCACGCTGCTC	19	18	150
TAS1R3	rs307377	C2269T	CTGGCCTTTCTCTGCTTCCT	CAAAGGAGACCCAGGTGATG	20	20	119
TAS1R3	rs307355	-C1572T	CGTGTGTGCTGTGAGCGTA	AATATGGCGCACATGCGAA	19	19	520
TAS1R3	rs35744813	-C1266T					
TAS2R16	rs846664	G516T	GGCTGAGGTGGAGAATTTTG	CCAGGAACAGGATGAAAGGA	20	20	231
TAS2R38	rs713598	C145G	CAATGCCTTCGTTTTCTTGGTG	GATGGCTTGGTAGCTGTGGT	22	20	190
TAS2R38	rs1726866	C785T	CCCACATTAAAGCCCTCAAG	TCTCCTCAACTTGGCATTGC	20	20	197
TAS2R38	rs10246939	G886A					

2.3. Optimization of the multiplex

In order to ensure the minimal primer dimer formation and to improve the amplification reaction performance two independent multiplexes were constructed joining the pairs of primers that better fit together. All the amplification reactions were executed in 2720 Thermal Cycler (Applied Biosystems) and Thermal Cycler (BioRad).

Multiplex 1

Multiplex 1 (M1) included the 5 SNPs selected from *TAS1R3* gene, rs34160967 from *TAS1R1* and rs713598 from *TAS2R38* gene. Before amplification reaction, a mixture containing all the forward and reverse primers has been prepared so that all of them had a final concentration of 2.0 µM in the mix. The final concentrations and volumes used in each PCR reaction are present in table 3.

Table 3 – Volumes added in one PCR reaction of Multiplex 1 for Portugal, Uganda, Mozambique, Angola and Equatorial Guinea samples.

Reagents	1 reaction
MyTaq™ HS Mix (Bioline)	5 µL
Primers Mix (M1)	1 µL
DNA	1 µL*
Deionized water	till 10 µL

*In some cases, such as Angola samples, larger volumes of DNA were used, always performing a final volume of 10 µL.

The multiplex PCR amplification program used is described in figure 6.

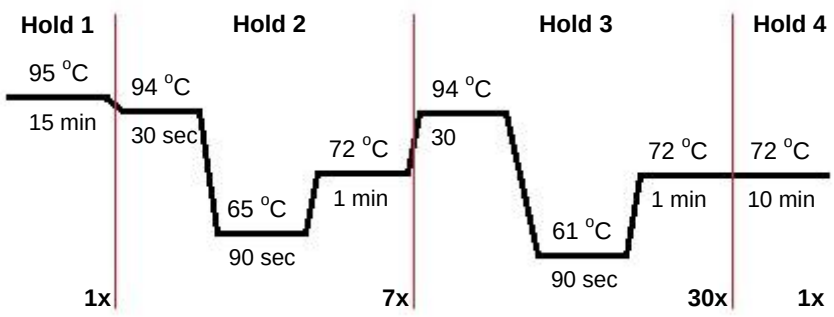


Fig. 6 – PCR program used for Multiplex 1.

Multiplex 2

Multiplex 2 (M2) was constituted by the remaining SNPs - rs41278020 and rs846664 from *TAS1R1* and *TAS2R16* genes, respectively and rs1726866 and rs10246939 belonging to *TAS2R38* gene. As in Multiplex 1, a previous mixture comprising forward and reverse primers has been done. All the primers were in the same concentration (2.0 µM) before being added to the PCR reaction. The volumes required to one PCR reaction are presented in table 4.

Table 4 – Volumes needed in one PCR reaction of Multiplex 2 for Portugal, Uganda, Mozambique, Angola and Equatorial Guinea samples.

Reagents	1 reaction
Qiagen® PCR Multiplex Kit	5 µL
Primers Mix (M2)	1 µL
DNA	1 µL*
Deionized Water	till 10 µL

*In some cases, such as Angola samples, larger volumes of DNA were used, always performing a final volume of 10 µL.

To carry out the PCR amplification for M2, the program presented in figure 7 was used.

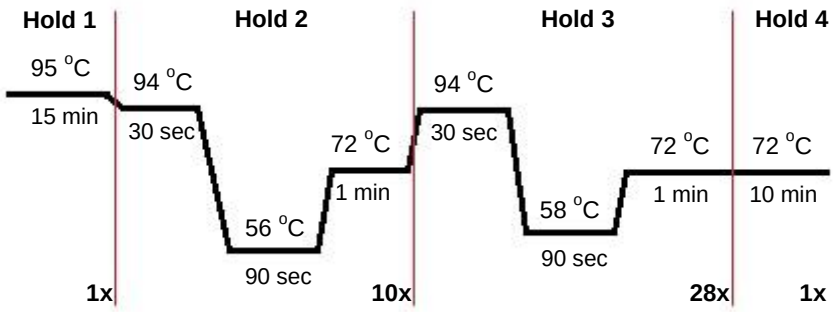


Fig. 7 – PCR program correspondent to Multiplex 2.

2.4. Electrophoresis

To detect the amplified fragments and to test for the possible existence of contamination, a polyacrylamide gel (T9C5) electrophoretic run was performed for each sample. The gel was stained following the Silver Staining standard procedure (figure 8).

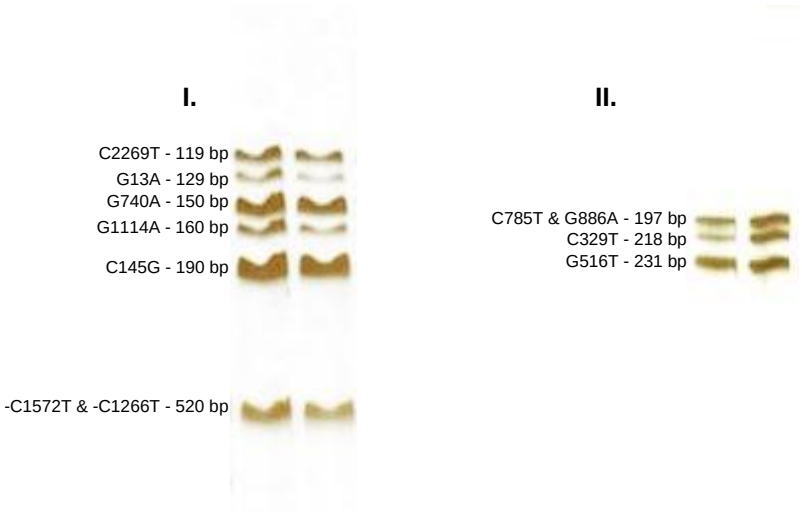


Fig. 8 – Band patterns observed after the electrophoresis for Multiplex 1 (**I**) and Multiplex 2 (**II**) and respective fragments length and SNPs.

3. SNP genotyping

The genotypic data from the eleven SNPs in study were accessed through minisequencing technique. The steps followed are described in the next points.

3.1. Single Base Extension Multiplexes

Eleven extension primers were designed to perform the minisequencing reactions, following the steps described to amplification primers and having the selected polymorphisms as targets.

Primer3 v.0.4.0 software (Rozen and Skaletsky, 2000) was used to design the extension primers, but in this case, it had the main role of evaluate the better option between the two possibilities, forward and reverse. Some parameters values were modified to better fit in the requirements of minisequencing reaction. The single base extension primers and their traits are present in table 5.

Table 5 – Single Base Extension primers.

Gene	SNP	Allele	Extension Primer	Length	Final length	Detection	Mutation
TAS1R1	rs41278020	C329T	TATGATGTGTGTTCTGACTCTG	22	22	C/T	C/T
TAS1R1	rs34160967	G1114A	AATGCCAAGCTTTCATG	17	17	G/A	G/A
TAS1R3	rs76755863	G13A	CTGAGGCCCAGGACAG	16	28	C/T	G/A
TAS1R3	rs111615792	G740A	CCTGGTGCCGCTGCCCC	17	34	G/A	G/A
TAS1R3	rs307377	C2269T	CGGGCACGGTTGTAGC	16	40	A/G	T/C
TAS1R3	rs307355	-C1572T	ACATGGTACACGCAAAGC	18	58	A/G	T/C
TAS1R3	rs35744813	-C1266T	CGGCACACATGCATGCC	17	52	A/G	T/C
TAS2R16	rs846664	G516T	TGGAAC**T**GATACTGATGAAA	20	70	A/C	T/G
TAS2R38	rs713598	C145G	GGATGTAGTGAAGAGGCAG	19	64	G/C	G/C
TAS2R38	rs1726866	C785T	CTTTGTGATATCATCCTGTG	20	76	C/T	C/T
TAS2R38	rs10246939	G886A	CTCTGGGCATGCAGCC	16	46	A/G	A/G

Note: The nucleotide in red (in *TAS2R16*) corresponds to a SNP present in the annealing local of extension primers.

3.2. Minisequencing protocol

After the multiplex amplification an initial purification with ExoSAP-IT®, in order to remove dNTPs and primers that have not been consumed and that could interfere in the minisequencing reaction was performed. Each reaction includes 1µL of ExoSAP-IT® (USB® Products; Affymetrix) and 1 µL of the amplified product, succeeding the digestion in a thermocycler at 37^{0}C for 15 minutes and 85^{0}C another 15 minutes.

The next step consists in annealing the specific probe in the target local and consequent addition of the fluorescent ddNTP, leading to the detection of the desirable base (Butler, 2005).

Two independent SNaPshot™ reactions were performed, one for Multiplex 1 and another for Multiplex 2, each one containing distinct extension primers. In tables 6 and 7 are summarized the reagents and extension primers specific for each reaction and the correspondent volumes necessary *per* reaction. The extension primers, in both SNaPshot™ reactions, had an initial concentration of 20 µM, presenting a different final concentration in accordance with the volume added to each reaction. The SNaPshot™ Multiplex Kit, provided by Applied Biosystems, comprises ddNTPs marked with different fluorescent dyes.

Table 6 – Components and respective volumes needed in one SNaPshot reaction for Multiplex 1.

Reagents/ SNP genotyped	1 reaction (µL)
G1114A	0.05
G13A	0.5
G740A	0.05
C2269T	0.05
-C1572T	0.8
-C1266T	0.1
C145G	0.1
SNaPshot™ Multiplex kit (ABI Prism®)	1.0
Purified PCR Product	2.0
Deionized Water	Till 5.0

Table 7 – Components and volumes required in one SNaPshot reaction for Multiplex 2.

Reagents/ Extension Primers	1 reaction (µL)
C329T	0.2
G516T	0.2
C785T	0.7
G886A	0.2
SNaPshot™ Multiplex kit (ABI Prism®)	1.0
Purified PCR Product	2.0
Deionized Water	Till 5.0

In the SNaPshot™ reaction, the purified PCR product plus the mix containing the extension primers, deionized water and SNaPshot™ Multiplex kit are submitted to three different conditions: 10 seconds at 96ºC, 5 seconds at 50ºC and 30 seconds at 60ºC, through 25 cycles.

Finally the SNaPshot reaction products are purified with SAP (USB® products) enzyme responsible for removing unincorporated fluorescent ddNTPs. The protocol followed is composed by two consecutive steps: 60 minutes at 37^{0}C and 15 minutes at 85^{0}C. To the product resulting from SNaPshot reaction was added 1 µL of SAP, making a final volume of 6 µL.

The SNaPshot™ reactions and purifications were carried out by 2720 Thermal Cycler (Applied Biosystems) and Thermal Cycler (BioRad).

The samples genotyping was performed in ABI 3130 Genetic Analyser (Applied Biosystems), after doing a previous mix with 12 µL Hi-Di™ Formamide (Applied Biosystems) plus GeneScan™-120 LIZ™ and 1 µL of final product. The formamide function is to keep the DNA in single chain and Liz 120 is the size standard. In figure 9 are represented the resulting electropherograms from GeneMapper® software (Applied Biosystems).

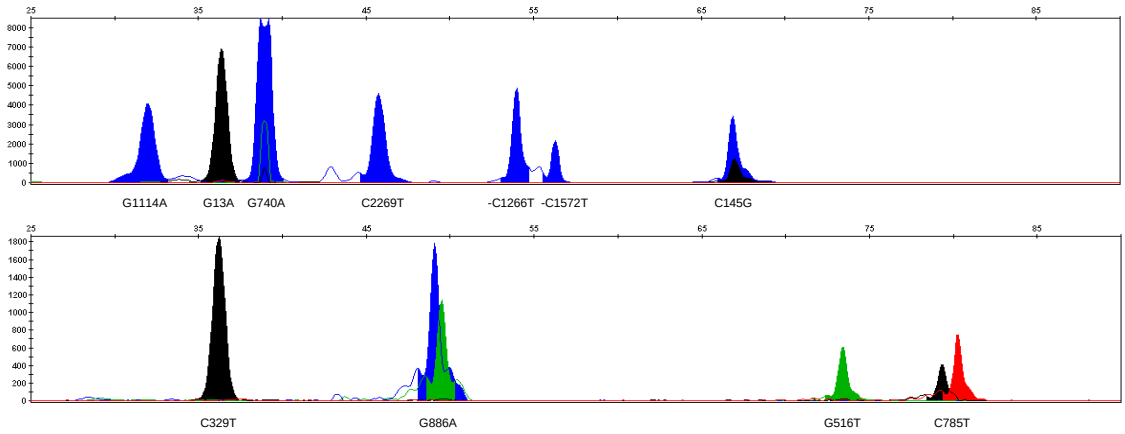


Fig. 9 – Electropherograms from Multiplex 1 and Multiplex 2, respectively. In the case of polymorphisms G13A, C2269T, -C1572T, -C1266T and G516T were designed reverse extension primers, which led to the detection of the complementary base.

A graphical overview of the steps followed in the protocols presented, since amplification, SNP genotyping and their analysis, are schematized in figure 10.

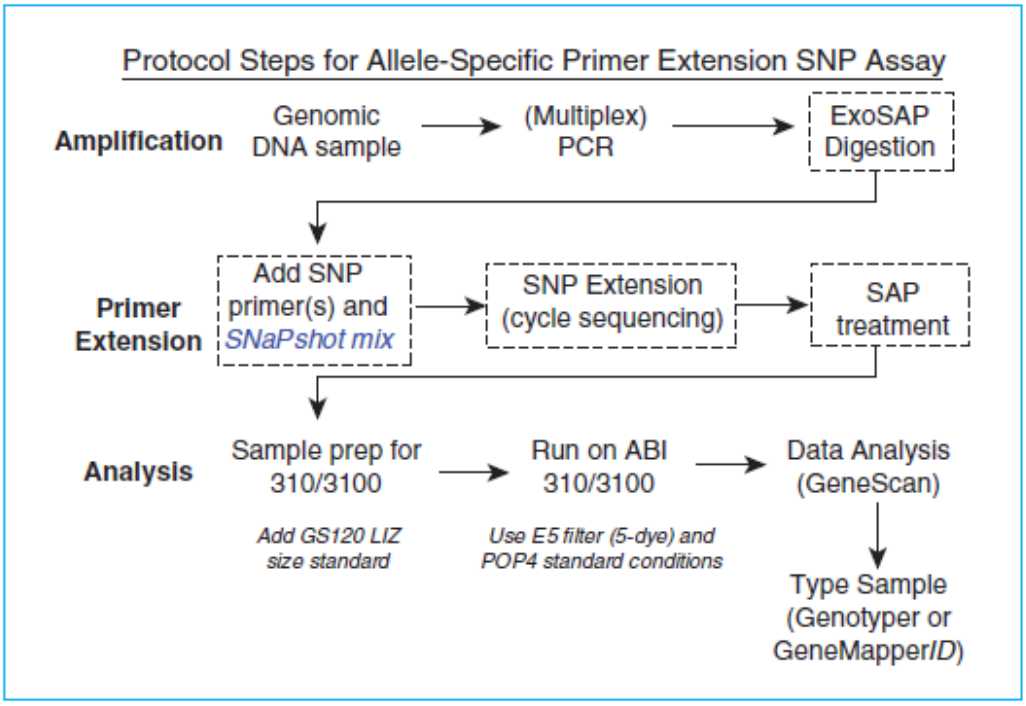


Fig. 10 – Procedures for allele-specific Primer extension SNP assay. The steps contained in boxes are the correspondent to SNP genotyping. Adapted from Butler, 2005.

4. Sanger Sequencing

To validate the results obtained in PCR and Single Base Extension reactions Sanger sequencing technique was carried out.

Similarly to minisequencing procedure, it was performed an initial purification with ExoSAP-IT®, used to eliminate the remaining dNTPs and primers that have not been consumed in the amplification reaction. The thermocycler program performed was the described in the protocol referred.

The sequencing reaction was prepared joining to each purified product 1 µL of the amplification primer (forward or reverse), sequencing buffer (1 µL) and sequencing kit (1 µL), and performing a final volume of 5 µL. The thermocycler program used to perform this reaction is described in table 8.

Table 8 – PCR program used in Sanger Sequencing.

Temperature	Duration	Number of cycles
96ºC	4 minutes	1
96ºC	10 seconds	
55ºC	5 seconds	35
60ºC	4 minutes	
60ºC	10 minutes	1

The sequencing reactions were done in 2720 Thermal Cycler (Applied Biosystems) or Thermal Cycler (BioRad).

After the sequencing reaction the products were purified through a Sephadex® column during 4 minutes at 4400 rpm, after an earlier centrifugation of 750 µL Sephadex® to form the columns. This purification is used to remove primers and ddNTPs, which are retained in the Sephadex® column. The products were then resuspended in 12 µL of Hi-Di™ Formamide and capillary electrophoresis was performed in ABI 3130 Genetic Analyser (Applied Biosystems).

5. Data analysis

SNaPshot results were analysed in GeneMapper® Software v4.0 and Peak Scanner™ software v1.0, both afforded by Applied Biosystems and the sequencing data were analised with Sequence Scanner™ v1.0 and Geneious Pro 5.5.6 softwares.

The Arlenquin software v3.5.1.2 (Excoffier and Lischer, 2010) was used to:

- Define the possible haplotypes for *TAS1R1*, *TAS1R3* and *TAS2R38*;
- Obtain estimates of allele and haplotype frequencies;
- Calculate F_{ST} values;
- Perform AMOVA;
- Test for Hardy-Weinberg Equilibrium.

The significance level assumed for multiple tests was corrected according to Bonferroni.

SPSS v20 software (SPSS Inc.) was used to obtain multidimensional scaling plot (MDS) representations, following the algorithm PROXSCAL, based on the F_{ST} values.

The contour map of derived allele T516 from *TAS2R16*, according to geography, was executed with Surfer v8.05 (Golden Software). To better visualize the relations of the haplotypes at *TAS1R1* and *TAS1R3* genes from the populations typed in the present work and other available in the literature, networks were constructed in NETWORK software v4.6.1.0 (Bandelt *et al.*, 1999).

For the contextualization of the genetic patterns observed in the studied populations, additional data were retrieved from the literature and databases available online. Data for the eleven SNPs were extracted from International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) (The International HapMap Consortium, 2010). With respect to G516T, other database, ALFRED (http://alfred.med.yale.edu/) (Rajeevan *et al.*, 2012) was consulted to obtain complementary information.

IV. RESULTS & DISCUSSION

1. *Locus* by *locus* approach

From the analysis of expected and observed heterozygosity, no deviations to Hardy-Weinberg Equilibrium were detected among our populations after the application of the Bonferroni correction for multiple tests (supplementary table 1).

1.1. Bitter taste

To access the genetic variation that influences bitter perception in human populations, two genes were selected – *TAS2R16* and *TAS2R38* – because, before, both had been clearly associated to the phenotypic variability in the discernment of bitter-tasting compounds.

TAS2R16

As previous referred, *TAS2R16* codes for a receptor mediating response to various β–glucopyranosides commonly found in nature that elicit a bitter taste. Within this gene only the SNP G516T (rs846664) was selected to be here screened because previous transient transfection studies have demonstrated that the two alleles at this SNP confer different ability to taste several glycosides compounds: the ancestral G516 allele was linked to lower sensitivity to those compounds comparing to the derived T516 allele (Soranzo *et al.*, 2005). For simplicity, G516 will be referred to as the non-taster allele whilst T516 as the taster one.

Estimates of T516 allele frequencies for the populations studied are presented in table 9.

Table 9 – *TAS2R16**T516 allele frequency plus standard deviation in different populations.

Sample / Allele	Uganda (n=49)	Angola (n=27)	Mozambique (n=34)	Equatorial Guinea (n=76)	Portugal (n=49)
T	0.8571±0.0355	0.6034±0.0648	0.7059±0.0557	0.7368±0.0358	1.0000±0.0000

In the four African populations, the ancestral G516 allele was usually found at much lower frequencies than the derived T516 and it was not detected in the individuals typed from Portugal. These results are in accordance with data previously reported for other populations revealing that the frequency of the taster allele, T516, is

1 or very near fixation outside Africa (Soranzo *et al.*, 2005). Apart from Africa, G516 was only rarely found in Middle Eastern populations or other with recognized African ancestry, such as African Americans (Hinrichs *et al.*, 2006).

Concerning the African populations here studied, all three Bantu-speaking/ agrarian populations – from Angola, Mozambique and Equatorial Guinea – showed higher frequencies of the non-taster G516 allele comparatively to the pastoralist society from Uganda. This allele reached its maximum frequency in Angola, being present in ~40% of Angolan chromosomes.

In order to put our results in a broader framework we recruited data available from the literature (Soranzo *et al.*, 2005; Hinrichs *et al.*, 2006) and from the HapMap (http://hapmap.ncbi.nlm.nih.gov/) (The International HapMap Consortium, 2010) or ALFRED (http://alfred.med.yale.edu/) (Rajeevan *et al.*, 2012) databases (supplementary table 3). A total of 76 populations were used in this comparative analysis. For G516, values of identical magnitude to that estimated for Angola were only registered in the Yoruba (36.7%) and the Ibo (42.4%), two agriculturalist groups from Nigeria, or in the Lisongo (38%), a group from Democratic Republic of the Congo.

As for the distribution of the derived taster T516, the highest frequency in African populations here tested was detected in the pastoralist population from Uganda (85.7%). In Africa, very elevated values (up to 75%) for this allele have been also reported in other pastoralist groups (the Mozabite from Algeria, the Maasai from Kenya or the Bedouin from North Africa), in hunter-gatherers (the San and the Biaka and Mbuti pygmies), and some agriculturalist populations (for instance Amhara and Zaramo, from Ethiopia and Tanzania, respectively). However, frequencies lower than 75% in Africa were only observed in agrarian groups.

Pairwise F_{ST} values were calculated between the populations typed in this study yielding the values presented in table 10.

Table 10 – F_{ST} values among Uganda, Angola, Mozambique, Equatorial Guinea and Portugal for *TAS2R16* SNP.

	Mozambique	**Uganda**	**Angola**	**Equatorial Guinea**	**Portugal**
Mozambique	*				
Uganda	0.0644	*			
Angola	0.0155	**0.1593**	*		
Equatorial Guinea	0.0000	0.0420	0.0299	*	
Portugal	**0.3372**	**0.1423**	**0.4703**	**0.2289**	*

The significant values after Bonferroni's correction are in blue.

Significant genetic distances were found between the 3 African agrarian populations and Uganda, as well as between all African populations and Portugal. Yet,

when the Bonferroni correction for multiple tests was applied, concerning African populations only the distance between Uganda and Angola remained significant.

Differences between Portuguese and Africans (0.1423 to 0.470) were in general considerably larger then between pairs of African populations (0.0000 to 0.1593), and the largest F_{ST} values observed were between Portugal and the agrarian African societies, especially from Angola (0.4703) and Mozambique (0.3372). The population from which Portugal differed less was the pastoralist from Uganda (0.1423).

Genetic distances were also obtained considering not only the populations here studied but in addition those listed in supplementary table 3; the corresponding F_{ST} values are shown in supplementary table 6.

To obtain a visual representation of the relationships between all populations, the matrix of F_{ST} distance was used to build a MDS plot which is present in figure 11.

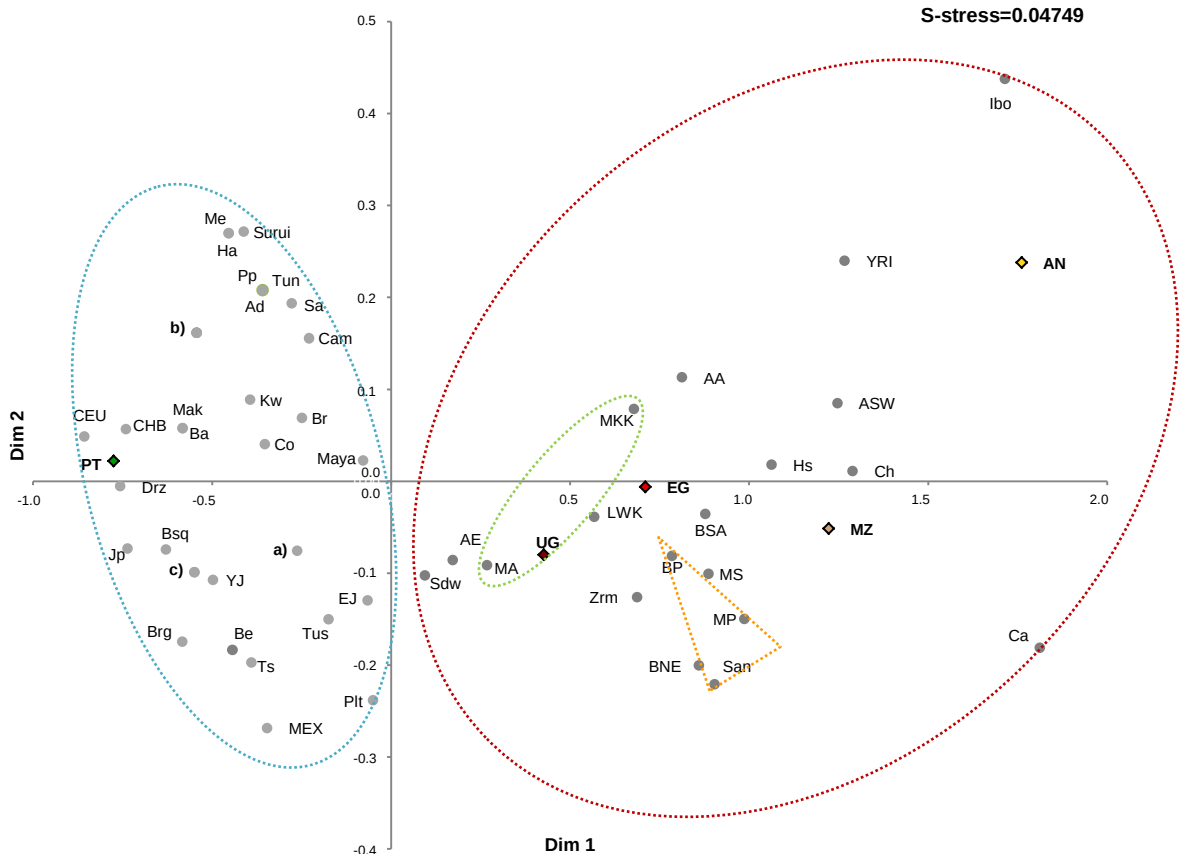


Fig. 11 – Multidimensional scaling plot of F_{ST} values corresponding to *TAS2R16* SNP. AN: Angola; EG: Equatorial Guinea; MZ: Mozambique; UG: Uganda; PT: Portugal; Ca: Cameroon; Li: Lisongo; YRI: Yoruba; Ch: Chagga; ASW: African ancestry in Southwest USA; Hs: Hausa; MP: Mbuti Pygmies; MS: Mandenka ; BSA: Bantu from South Africa; BNE: Bantu from North-eastern Africa; BP: Biaka Pygmies; AA: African-American; Zrm: Zaramo; EG: Equatorial Guinea; MKK: Maasai; LWK: Luhya; MA: Mozabite; AE: Amhara; Sdw: Sandawe; EJ: Ethiopian Jews; Plt: Palestinian; Br: Brahui; Tus: Tuscan; Cam: Cambodian; Sa: Sardinian; Tun: Tunisian; MEX: Mexican; Ts: Tsaatan; Co: Colombian; Kw: Kuwaiti; Ad: Adygei; Pp: Papuan; Ha: Hazara; Orc: Orcadian; YJ: Yemenite Jews; Be: Bedouin; Bsq: Basque; Brg: Bergamo; Drz: Druze; Mak: Makrani; JPT: Japanese; CEU: Utah residents with European ancestry. Xibo, Mongolian, Tu, Miaozu, Yizu, She, Tujia, Hezhen, Dai, Lahu, Naxi, Daur, Oroqen and Uygur are represented by a) and b) contains Pathan, Burusho, Kalash, Russian, Pima and Yakut populations. In c) are represented Sindhi, French and Karitiana individuals.

As can be seen, the 75 worldwide populations, encompassing all continents, are grouped in two main clusters – one with the majority of African populations (red circle) and other with all the Non-African populations plus a few of North-African ones such as Bedouin and Tunisian (blue circle).

Figure 11 also reveals that non-African populations are quite concentrated in their cluster, whereas the African ones stand much more dispersed between each other. By other words, this means that diversity for G516T is considerably reduced out of Africa compared to that registered inside the African continent.

Within the African cluster, the sample from Uganda, the Maasai from Kenya (MKK) and the Mozabite from Algeria, all constituting pastoralist groups, are positioned in the periphery of the cluster (green oval), quite near the group of non-African populations. Hunter-gatherers, such as the San or the Biaka and Mbuti Pygmies, also assemble near each other (orange triangle). So, despite not being clearly differentiated from the remaining African populations, either pastoralists or hunter-gatherers tend to cluster with populations with identical lifestyle. African agriculturalist populations are much more widely dispersed in the African cluster, but, interestingly, Bantu-speaker groups mainly from Western Sub-Saharan Africa (samples from Angola and Cameroon, or the Ibo) occupy in the MDS plot another peripheral side on the cluster of African populations. Angola, Cameroon and the Ibo are indeed the populations among which the frequency of the non-taster G516 allele systematically reaches the highest values.

In overall, the clustering pattern of the African populations, although indicating that the distribution of G516T in Africa might be in a certain extent related with the typical lifestyle of populations, suggests that geography is a variable clearly influencing the diversity at this polymorphism.

In figure 12 is illustrated the contour map across Africa, Europe and Asia of T516 allele frequency.

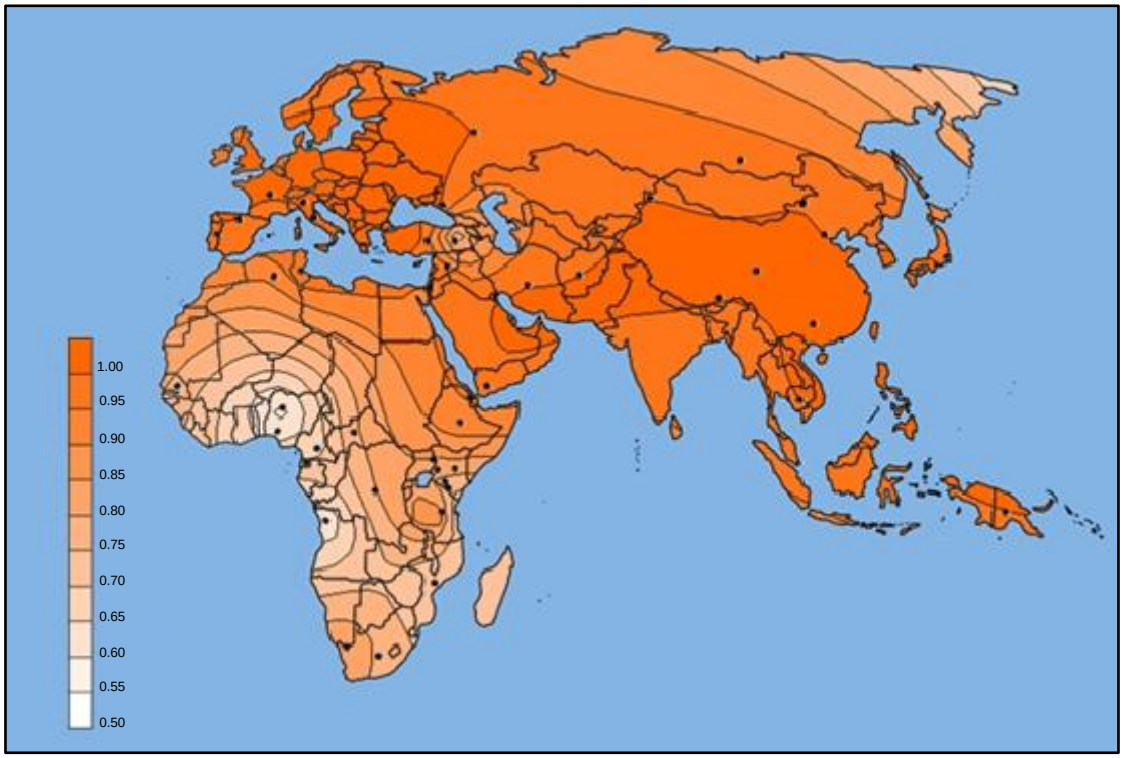


Fig. 12 – T516 frequency contour map of Africa and Eurasia. The remaining continents (Oceania and America) were omitted due the lack of enough data.

This distribution is quite intriguing since in the African continent it can be seen a remarkable correspondence between the frequency distribution of this SNP and the Bantu dispersion. In fact, the region showing the lowest frequencies of T516 is rather coincident with the centre of origin of Bantu languages (Diamond and Bellwood, 2003; Jobling *et al.*, 2004) in Cameroon and Eastern Nigeria. From that region, the frequency of the taster allele tends to increase, suggesting a dispersion pattern that evokes the Bantu dispersal routes. Since agriculturalist lifestyle is intrinsically correlated with Bantu people, the strong representativeness of the ancestral allele in these populations can be related to their dietary habits, possibly indicating that the detection of glycosides compounds to agrarian groups may not be as relevant as to pastoralists and hunter-gatherers.

To further dissect the factors contributing to explain the pattern of distribution of G516T, several AMOVA tests were performed using different criteria to establish groups of populations.

Firstly we have restricted the analysis to African populations and grouped them according to lifestyle, geography and malaria risk. This latter assay was tested because previously, Soranzo *et al.* (2005), based on a certain similarity in the distribution of the low-sensitivity allele and that of some malaria-resistant alleles, speculates that the ancestral G516 allele could confer a modest selective advantage in malaria-infested

areas through favouring chronic low level ingestion of cyanogenic glycosides. Although some β-glucopyranosides compounds are toxic to humans, their moderate consume has been associated with an increased protection against malaria (Soranzo *et al.*, 2005; Jackson, 1990).

Information for classifying populations based on malaria risk was recruited from http://cdc-malaria.ncsa.uiuc.edu/.

In the three AMOVA scenarios tested, the major proportion of variation was always ascertained to the within populations component, which was never less than ~93% of the total variation observed for G516T (table 11).

Table 11 – AMOVA results in Africa.

Source of Variation	Percentage of variation	*P*-value
Lifestyle		
Among groups	3.11	0.0287±0.0017
Among populations within groups	3.80	0.0000±0.0000
Within populations	93.09	0.0000±0.0000
Geography (by region)		
Among groups	3.21	0.0085±0.0009
Among populations within groups	2.09	0.0002±0.0001
Within populations	94.70	0.0000±0.0000
Malaria Distribution		
Among groups	4.12	0.0045±0.0007
Among populations within groups	2.95	0.0000±0.0000
Within populations	92.93	0.0000±0.0000

However, the proportions of variation among groups, although being incomparably much smaller than those due to differences within populations, were also always significant in three distinct scenarios – lifestyle, geography and malaria distribution. When lifestyle was considered (hunting-gathering, agriculture and pastoralism) the percentage of variation among groups was 3.11%. That proportion increased to 3.21% when the geographical criterion was applied (the groups were: Central Africa, East Africa, West Africa, South Africa, North Africa and South-East Africa). But the variation among groups reached the highest value, 4.12%, when the malaria risk was used in the clustering system.

Taken together these results indicates that, from the 3 factors tested, malaria risk contributes more to explain the distribution of G516T in Africa than geography, and

both malaria risk and geography more than lifestyle. Since malaria distribution in Africa is obviously correlated with geography, the meaning of the differences in the AMOVA tests need to be interpreted with much caution.

Next, AMOVA was performed considering the entire set of populations available across the world using the same three criteria for clustering populations, but additionally testing African versus non-African populations when geography was assumed. Results are presented in table 12.

Table 12 – AMOVA results corresponding to groups formed with data from 76 populations with different regions and ethnics.

Source of Variation	Percentage of variation	*P*-value
Lifestyle		
Among groups	1.69	0.1426±0.0038
Among populations within groups	7.73	0.0000±0.0000
Within populations	90.58	0.0000±0.0000
Geography (by region)		
Among groups	14.49	0.0000± 0.0000
Among populations within groups	2.20	0.0000± 0.0000
Within populations	83.31	0.0000±0.0000
Malaria Distribution		
Among groups	7.76	0.0002±0.0001
Among populations within groups	10.34	0.0000±0.0000
Within populations	81.90	0.0000±0.0000
Africa vs. Non-Africa		
Among groups	15.43	0.0000±0.0000
Among populations within groups	6.86	0.0000±0.0000
Within populations	77.72	0.0000±0.0000

At a worldwide level, differences among populations determined by lifestyle decreased to 1.69% and lost statistical significance.

Contrarily, the proportions of variation ascribed to geography or malaria risk substantially increased, reaching very highly significant values. However, whereas malaria distribution accounted to explain 7.8% of the total variation at G516T, geography explained approximately two times more differences among groups: 14.49%

when the region of origin was considered and even more, 15.43% when Africans were contrasted to non-Africans.

The recognition of bitter natural toxins is thought to may have conferred an important selective advantage during human evolution, alerting humans to noxious foods (Soranzo *et al.*, 2005), having been admitted that in hunter-gatherers' communities, sensitivity to bitterness was presumably advantageous because of various noxious plants (Li *et al.*, 2011). Both Soranzo *et al.* (2005) and later Li *et al.* (2011) reported on signs of selection at *TAS1R16*, with the first authors having further hypothesized that the global pattern at this gene has resulted from a balance between protection against malaria and protection against toxins in malaria-free zones.

Our findings indicate that whereas in Africa the distribution of G516T is only faintly correlated with lifestyle, demographic factors, likely related with the Bantu expansion, might have played an important role in shaping diversity at this SNP. In fact, the excess of the non-taster G516 allele in Western regions of Africa affected by malaria, which Soranzo *et al.* (2005) interpreted as a sign of selection driven by malaria, seems as well be explainable by the effects of the Bantu expansion.

For this polymorphism differences between Africans and non-Africans, although statistically significant, are not higher than usually registered, indicating that bottleneck effects consequent to the out of Africa dispersion of humans seem enough to explain the differences observed.

TAS2R38

The three SNPs screened in *TAS2R38* define different haplotypes that have been associated with the phenotypic variation in PTC perception, from which the two most common are PAV (proline, alanine, valine) and AVI (alanine, valine, isoleucine). PAV is considered to be a dominant "taster" haplotype and AVI a recessive "non-taster" one, since they were proven to be strongly correlated with PTC bitter taste sensitivity and insensitivity, respectively.

In table 13 are presented the haplotype frequencies estimated in the 5 populations studied in this work.

Table 13 – *TAS2R38* haplotypes frequencies and diversities.

Sample / Haplotype	Uganda (n=45)	Angola (n=17)	Mozambique (n=23)	Equatorial Guinea (n=76)	Portugal (n=49)
PAV	0.4387±0.0543	0.4018±0.0884	0.4440±0.0754	0.5197±0.0414	0.5091±0.0511
PVV	0.0168±0.0160	0.0356±0.0428	0.0336±0.0325	-	-
PAI	-	0.0321±0.0375	0.0224±0.0217	-	-
PVI	-	0.0599±0.0473	-	-	0.0215±0.0147
AAV	0.0178±0.0158	0.0626±0.0461	0.0227±0.0217	0.0066±0.0064	0.0419±0.0200
AVV	0.0156±0.0149	-	-	-	-
AAI	0.2435±0.0471	0.2093±0.0795	0.2721±0.0721	0.1776±0.0303	0.0204±0.0148
AVI	0.2676±0.0477	0.1986±0.0786	0.2055±0.0643	0.2961±0.0364	0.4071±0.0512
Haplotype Diversity	0.6879±0.0248	0.7594±0.0496	0.7053±0.0398	0.6147±0.0235	0.5765±0.0268
Non-tasters	0.4387	0.4018	0.4440	0.5197	0.5091
Tasters	0.2676	0.1986	0.2055	0.2961	0.4071
Intermediate	0.2937	0.3996	0.3505	0.1842	0.0838

The most common haplotype in all populations was PAV, showing frequencies always higher than 40% and reaching up to 50% in Portugal and Equatorial Guinea. The second most common haplotype in Portugal was the non-taster AVI, being present with the high value of 41%. AVI is also well represented in the African populations, but ranging only the intermediate values of 19.9% to 29.6%.

Besides PAV and AVI, other haplotypes were detected, commonly referred to as intermediate haplotypes. Their combined frequency in each population is also shown in table 13. In Portugal only three intermediate haplotypes were found occurring at the combined low frequency of 8.4%. Contrarily, in African populations the number and frequency of intermediate haplotypes were typically much higher. Among them, AAI

was systematically quite common in Africans (17.8-27.2%) being even more frequent than the non-taster haplotype in Angola and Mozambique. This haplotype was also detected in Portugal, although at much lower frequency (2%). Together AAI and AAV, constituted the two unique intermediate haplotypes shared by all the populations here examined.

Most of the intermediate haplotypes detected in this study had been already described in other populations from different geographical regions, although only attaining high levels of diversity in Sub-Saharan African groups (Wooding *et al.*, 2004; Wooding *et al.*, 2010; Campbell *et al.*, 2012).

However, two of these haplotypes had not been reported before, leading to presume that they were firstly observed in this study. These two haplotypes were:

- **PVV**, detected in Uganda, Angola and Mozambique;
- **AVV**, found to be present only in Uganda.

In table 14 are presented the F_{ST} values obtained for the pairs of populations here studied.

Table 14 – F_{ST} values based on *TAS2R38* haplotype frequencies.

	Angola	**Mozambique**	**Equatorial Guinea**	**Uganda**	**Portugal**
Angola	*				
Mozambique	0.0000	*			
Equatorial Guinea	0.0222	0.0280	*		
Uganda	0.0056	0.0068	0.0197	*	
Portugal	0.0667	**0.0949**	0.0415	**0.0663**	*

The significant values after Bonferroni's correction are in blue.

No significant genetic differentiations were found between the African populations, but all of them differed from Portugal, although applying the Bonferroni correction for multiple tests only the genetic distances between Portugal/Mozambique and Portugal/Uganda remained significant.

Comparisons were then extended to other populations whose data was collected from literature (supplementary table 7). The obtained F_{ST} values, shown in supplementary table 9, were used to construct the MDS plot displayed in figure 13.

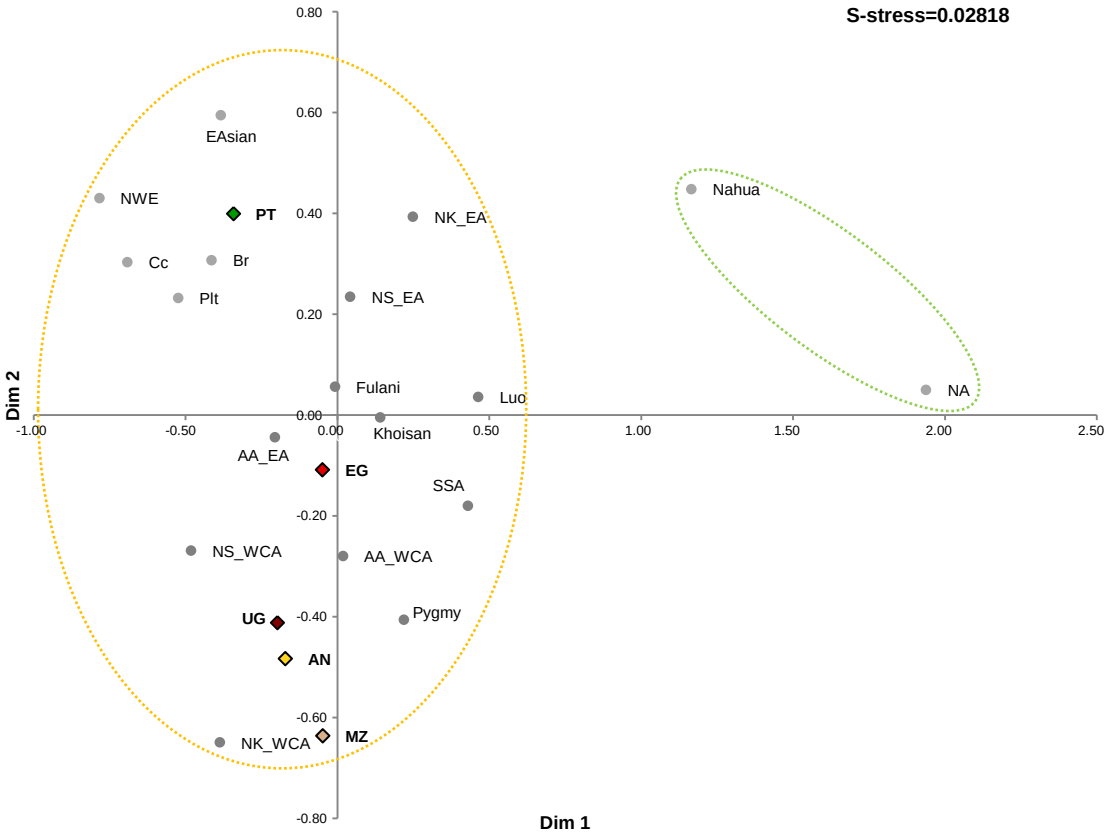


Fig. 13 – Multidimensional Scaling of F_{ST} values corresponding to *TAS2R38* haplotypes. AN: Angola; EG: Equatorial Guinea; MZ: Mozambique; UG: Uganda; PT: Portugal; NA: Native American; NWE: North-Western European; EAsian: East Asian; Cc: Caucasian; NK_EA: Niger-Kordofanian from East Africa; NS_EA: Nilo-Saharan from East Africa; AA_EA: Afro-asiatic from East Africa; NK_WCA: Niger-Kordofanian from West Central Africa; NS_WCA: Nilo-Saharan from West Central Africa; AA_WCA: Afro-asiatic from West Central Africa; SSA: Sub-Saharan Africa.

As is evident from the analysis of this figure, there are 2 populations that emerge as clear outliers in the worldwide pattern of diversity of *TAS2R38*. They are the two samples of Native Americans. Concerning these populations, one sample is from a non-identified group from South-Western North America (Wooding *et al.*, 2004) and the other is from the Mexican Nahua, and both samples show highly significant distances with the remaining populations considered in the comparison. This strong differentiation is due to the dramatically high frequency of the PAV haplotype together with the absence of intermediate haplotypes in two Native American groups, a distribution pattern that likely resulted from strong bottleneck effects during the settlement of the Americas.

The Eurasian populations tend to cluster near each other in quadrant 2 of the MDS representation, standing slightly differentiated from the African populations (figure 13). What seems to account more to the positioning of Eurasian populations relatively to the African ones, is a) the absence or rarity of the intermediate haplotype AAI, which, by turn, is rather common in African populations (minimum value in Africans: 10%;

maximum value in Eurasians: 5%), and b) the usual high frequency of the non-taster AVI haplotype (up to 40% in European and the majority of Asian populations, down to 37% in Africans).

Despite that, average F_{ST} values between European and African populations are relatively low compared to estimates based on other genetic variations here examined (see, for instance *TAS1R16*) or on global comparisons between worldwide populations yielding F_{ST} values that typically range from 0.10 to 0.16 (Campbell and Tishkoff, 2008).

Concentrating now on the African populations, the level of differentiation between them is also notably low, as can be seen in the MDS plot of figure 13, revealing that the African populations are quite near each other. This reflects the finding that the common haplotypes were present at rather similar frequencies in diverse African populations regardless of their geographical location or lifestyle. As a matter of fact, representative of hunter-gatherers (Khoisan – Hadza and Sandawe - from East Africa and Pygmy from West Central Africa), pastoralists (Uganda, Luo and Fulani) and agriculturalists (Angola, Mozambique, Equatorial Guinea, Nilo-Saharan and Niger-Kordofanian from West Central Africa and Niger-Kordofanian from East Africa) are randomly spread in the cluster of African populations (figure 13) and exhibit some of lowest F_{ST} values in the entire matrix of genetic distances (supplementary table 9).

As a whole, these results seem to indicate that the *TAS2R38* haplotype variation in Africa is not associated with population lifestyle or diet.

For instance, three of the African populations here studied – Angola, Mozambique and Uganda – when contrasted with the sample of Niger-Kordofanian individuals from East Africa (Kenya) revealed genetic distances in the upper range of the values found in Africa. The 3 F_{ST} values were significant assuming the conventional P=0.05, although the significance was lost with the application of Bonferroni's correction. These Bantu-speakers from Kenya are farmers, as well are our sampled groups from Angola, Mozambique or Equatorial Guinea. So, sharing a mode of subsistence does not appear to imply increased genetic affinities for *TAS2R38*. Besides, geography within Sub-Saharan Africa also does not seem to explain the pattern of *TAS2R38* haplotype variation. The referred Kenyan group differs less from the group of Equatorial Guinea (0.0209), located in the Atlantic coast of Africa opposite to that near Kenya, than from its more neighbour group from Uganda (0.0587) (supplementary table 9).

Indeed, no signs arose that relationships between African populations regarding *TAS2R38* could be influenced in a certain extend by their geographical proximity.

For *TAS2R38*, we also conducted AMOVA. Firstly we used the entire set of populations, grouping them in Africans versus non-Africans. Under this criterion, the

percentage of variation ascertained to differences between groups only reached the value of 3.62%, which was just slightly higher than the proportion due to differences among populations within groups (table 15). When populations were grouped according to the continent of origin, variation between groups increased somewhat, but only to 5.19% while that among groups decreased to 0.89%. These results reinforce what figure 13 already brought up, that concerning *TAS2R38* haplotypes the level of genetic differentiation among global worldwide populations is rather low.

Table 15 – AMOVA results to *TAS2R38* corresponding to groups formed with the 23 populations represented in figure 13.

Source of Variation	Percentage of variation	*P*-value
Geography (by region)		
Among groups	2.79	0.0015±0.0004
Among populations within groups	0.86	0.0077±0.0008
Within populations	96.35	0.0000±0.0000
Geography (by continent)		
Among groups	5.19	0.0000±0.0000
Among populations within groups	0.89	0.0011± 0.0003
Within populations	93.92	0.0000±0.0000
Africa vs. non-Africa		
Among groups	3.62	0.0000±0.0000
Among populations within groups	1.69	0.0000±0.0000
Within populations	94.70	0.0000±0.0000

Additional AMOVA tests were performed using only African populations and grouping them according to lifestyle, language family and geographic region. However, regardless the way populations were grouped, no significant variation among groups was found (results not shown). This means that within ethnically diverse Africans with different diets/lifestyles, typically showing high levels of genetic substructure, and living in distinct regions of Africa, no major differences exist in the distribution of *TAS2R38* haplotypes, indicating that their variation is not correlated with diet, mode of subsistence or geographical environment.

On the whole, our results add support to some peculiar aspects of the genetic diversity distribution at the *TAS2R38 locus* that had been previously noticed (Campbell *et al.*, 2012). Firstly, it should be noted the remarkably similar frequencies of the common haplotypes in divergent African populations. Comparatively to populations

outside of Africa, African groups possess a wider range of *TAS2R38* haplotype diversity and consequently a wide range of bitter taste sensitivity (Campbell *et al.*, 2012). However, evidence up to now accumulated reveals that the distribution of such diversity is not due to local adaptation to diet.

Furthermore, the persistence in non-Africans of two of the most common haplotypes in Africans, accounting typically for more than 90% of *TAS2R38* diversity outside Africa, suggests not only that those haplotypes are functionally important, but also that they might had been maintained at high frequencies across worldwide populations in response to selective pressures. Results obtained to date are leading to exclude that such selective factor can be related with diet. However, given the striking pattern of distribution and other clear signs that *TAS2R38* is under selection, further studies are needed to explore the role of still unknown non-dietary biological processes in modelling diversity at this *locus* that mediates bitter taste perception.

1.2. *Umami* taste

As already mentioned in point 3.2 of chapter I (Introduction), *umami* taste is one of five basic taste qualities and plays a key role in the intake of amino acids. *Umami* is elicited by L-glutamate, typically as its Na salt (monosodium glutamate: MSG), some amino acids and purine nucleotides (such as IMP and GMP). It is known that taste sensitivity to *umami* substances varies widely among individuals. Distribution of individual MSG thresholds shows a bi-modal curve, and taste thresholds of MSG differ about 5-fold between taster and hypotaster groups (Lugaz *et al.*, 2002; Shigemura *et al.*, 2009b).

One well established *umami* receptor is a heterodimeric G protein-coupled receptor, consisting of the proteins T1R1 (taste receptor type 1, member 1) and the T1R3 (taste receptor type 1, member 3), which are encoded by *TAS1R1* and *TAS1R3* genes, respectively (Nelson *et al.*, 2002). Both *TAS1R1* and *TAS1R3* were studied in this work, using the information provided by the haplotypes defined by the SNPs screened in each gene. Due to the very high rate of unsuccessful genotyping in the sample from Angola, this sample was excluded from the *TAS1R1* and *TAS1R3* analyses.

TAS1R1

In this work we selected two common non-synonymous SNPs at *TAS1R1* showing evidence of being implied in *umami* perception: C329T and G1114A (Shigemura *et al.*, 2009b; Raliou *et al.*, 2009). Since the T allele at position 329 and the G at position 1114 were both associated with low sensitivity to glutamate, we assumed that the TG haplotype confers low *umami* perception, the CA haplotype high sensitivity, while the remaining two combinations are intermediate haplotypes (Raliou *et al.*, 2009).

Table 16 – *TAS1R1* haplotype frequencies and diversities.

Sample / Haplotype	Uganda (n=37)	Mozambique (n=32)	Equatorial Guinea (n=76)	Portugal (n=49)
CA	0.0541±0.0258	0.0156±0.0155	0.0592±0.0185	0.1122±0.0321
CG	0.9459±0.0258	0.9844±0.0155	0.9211±0.0209	0.8674±0.0341
TG	-	-	0.0197±0.0108	0.0204±0.0141
Haplotype Diversity	0.1037±0.0469	0.0312±0.0300	0.1487±0.0382	0.2371±0.0529

In table 16 are presented the *TAS1R1* haplotype frequencies estimated for Uganda, Mozambique, Equatorial Guinea and Portugal. Three different haplotypes were found, CA, CG and TG, with CG, the unique intermediate haplotype detected in the present work, being the overwhelmingly predominant haplotype in the four populations, occurring at its lowest frequency, 86.7%, in Portugal. The CA haplotype was also shared by all populations, but showed the highest frequency among the Portuguese, 11.2%. The TG combination was only rarely found in Equatorial Guinea and Portugal, accounting for around 2% of chromosomes in both populations.

For the *TAS1R1 locus*, Portugal presented greater haplotype diversity than any of the 3 African populations here studied, which is in opposition to the usual observation of highest levels of diversity among Africans. Nevertheless, haplotype diversities for all populations are visible moderate or low.

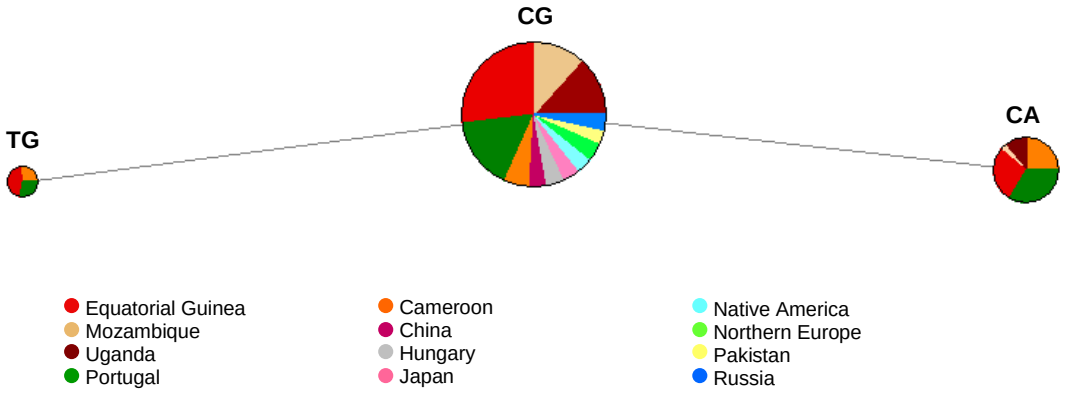


Fig. 14 – Network of *TAS1R1* haplotypes.

For *TAS1R1* few data was available for the comparisons. The more comprehensive study across the world was performed by Kim *et al.*, 2006, who have analysed a total of 8 populations, one from African (Cameroon), other from Native-American and the remaining from Eurasian regions. However the sizes of the 8 samples were particularly low (minimum 8, maximum 20 individuals *per* population), and so the study only provided rough estimates of haplotype frequencies for *TAS1R1*.

Even so, data from the study of Kim *et al*. (2006) and from our own work are globally concordant (figure 14). The 3 different haplotypes here identified were also the unique reported by Kim *et al.* (2006). These authors only detected the intermediate haplotype, CG, in 7 out the 8 studied populations, which is understandable given, at one hand, the extremely high frequency of this haplotype, and on the other hand, the very small number of individuals analysed. Likely, if samples sizes were enlarged, other haplotypes would be found, but probably at low or intermediate frequencies, as we did observed in this work. In the study of Kim *et al.* (2006), the 3 haplotypes were

exclusively detected in the sample from Cameroon, which was indeed the largest sample they screened, comprehending 20 Cameroonians. Among them, the CG haplotype represented 75% of the sampled chromosomes, constituting the lowest frequency out of the populations now available, while the combined frequency of the haplotypes, CA and TG, summed up to 25%. As a matter of fact, haplotype diversity in the Cameroonians was the highest reported to date, meaning that for *TAS1R1* Africans do not present necessarily lower diversities than non-Africans, as our data could indicate.

Pairwise F_{ST} distances between populations from our study and that of Kim *et al.* (2006) were calculated and are shown in table 17.

Table 17 – F_{ST} values correspondent to *TAS1R1* haplotypes. Mz: Mozambique; Ca: Cameroonian; EG: Equatorial Guinea; UG: Uganda; Ch: Chinese; NA: Native American; Hu: Hungarian; Pa: Pakistani; Ja: Japanese; Ru: Russian; PT: Portugal; NE: Northern European.

	MZ	**Ca**	**EG**	**UG**	**Ch**	**NA**	**Hu**	**Pa**	**Ja**	**Ru**	**PT**	**NE**
MZ	*											
Ca	**0.2252**	*										
EG	0.0465	**0.1280**	*									
UG	0.0367	0.1452	0.0241	*								
Ch	0.0071	0.1647	0.0402	0.0340	*							
NA	0.0071	0.1647	0.0402	0.0340	0.0303	*						
Hu	0.0071	0.1647	0.0402	0.0340	0.0303	0.0303	*					
Pa	0.0000	0.1498	0.0332	0.0262	0.0303	0.0303	0.0303	*				
Ja	0.0071	0.1647	0.0402	0.0340	0.0303	0.0303	0.0303	0.0303	*			
Ru	0.0071	0.1647	0.0402	0.0340	0.0303	0.0303	0.0303	0.0303	0.0303	*		
PT	0.0889	0.0514	0.0374	0.0459	0.0753	0.0753	0.0671	0.0671	0.0753	0.0753	*	
NE	0.0071	0.1647	0.0402	0.0340	0.0303	0.0303	0.0303	0.0303	0.0303	0.0303	0.0753	*

The significant values after Bonferroni's correction are in blue.

Obviously the distances between the populations from the work of Kim *et al.* (2006), where only the CG haplotype was detected, presented very similar values among them, but these results need to be cautiously interpreted given the limitation of the small samples size. Despite that, the remaining F_{ST} values were in general very low, and accordingly non-significant, excepting those referring to the distances involving the Cameroonians. The sample from Cameroon exhibited the greatest level of differentiation for *TAS1R1*, showing significant distances with almost all populations when the conventional P=0.05 was considered. When the correction for multiple tests

was applied, distances with Mozambique and Equatorial Guinea remained significant. The population from which the Cameroonians differed less was Portugal, with the low F_{ST} value of 0.0514 translating well the similarity of the distribution of the *TAS1R1* haplotypes in the two populations.

Kim *et al.* (2006) did not give details on the samples analysed, but their region of origin. Nevertheless, diversity at *TAS1R1* across populations is very narrow and does not seem to be correlated with diet/lifestyle. At least our sample of pastoralist from Uganda is not statistically different from the agriculturalists from Mozambique and Equatorial Guinea.

Geography also does not represent a factor appearing to influence the general distribution at the *loci*. When we conducted AMOVA assuming groups according to geography (continent of origin) no significant variation among groups was detected, which was predictable given the absence of sharp dissimilarities across populations.

In the future, more comprehensive studies are needed to obtain a better scenario of the distribution of the *TAS1R1* haplotypes at a worldwide level. However, its variation seems to be relatively uniform across human populations, indicating that *TAS1R1* does not contributes much to differences in *umami* taste sensitivity among them.

This uniformity in diverse populations, from distant geographical origins and different lifestyles, raises the questions of the factors that underlie such distribution pattern. Previous evolutionary analyses support that several genes belonging to the TAS1R family, including *TAS1R1*, has been under positive selection (Kim *et al.*, 2006). Yet, if selection has somehow operated in this gene, it must not have been driven by diet or geographical related pressures.

TAS1R3

With respect to *TAS1R3*, we have studied 3 non-synonymous variations for which there are indications that they might partially account for the inter-individual variability in *umami* taste perception: G13A, G740A and C2269T (Chen *et al.*, 2009; Raliou *et al.*, 2009; Shigemura *et al.*, 2009b). For this *locus*, we also inferred the haplotypes defined by the 3 tested SNPs. From the 3 selected SNPs, C2269T is the unique for which functional evidence was gathered revealing that the T allele is associated with diminished sensitivity to substances eliciting the *umami* taste (Raliou *et al.*, 2009, Chen *et al.*, 2009; Shigemura *et al.*, 2009b). For this reason, we have assumed that haplotypes harbouring C at this SNP were high sensitivity haplotypes while those carrying the T allele were low sensitivity haplotypes.

Table 18 – *TAS1R3 umami* haplotype frequencies and diversities.

Haplotype \ Sample	Uganda (n=47)	Mozambique (n=34)	Equatorial Guinea (n=76)	Portugal (n=49)
GAC	0.1170±0.0312	0.2794±0.0558	0.1974±0.0329	0.0102±0.0099
GGC	0.8724±0.0320	0.7059±0.0577	0.7829±0.0341	0.9286±0.0252
GGT	0.0106±0.0097	-	-	0.0612±0.0238
GAT	-	0.0147±0.0150	0.0197±0.0115	-
Haplotype Diversity	0.2276±0.0527	0.4298±0.0499	0.3500±0.0405	0.1353±0.0456

The haplotype frequencies estimated for *TAS1R3* in different populations are presented in table 18. Our screening resulted in the detection of four haplotypes, but not simultaneously present in all populations. The low sensitivity haplotype, AAT, was not observed. In all populations the clearly predominant haplotype was GGC, the haplotype associated with increased sensitivity, reaching the very high value of 92.9% in Portugal. This haplotype was also very well represented in Africans, but only ranging from 70.6% to 87.2%. In all African populations the second most common haplotype was GAC, supposedly also conferring increased *umami* taste sensibility, always occurring with frequencies up to 11%. Two other less-taster haplotypes were detected, although much more rarely observed: GGT, present in Uganda and Portugal, and GAT present in Mozambique and Equatorial Guinea, although accounting for less than 2% of chromosomes in both populations.

As commonly observed for the majority of *loci*, the 3 African populations from Uganda, Mozambique and Equatorial Guinea all revealed higher haplotype diversities at *TAS1R3* than the unique non-African population investigated in this work, the Portuguese sample.

Comparing our results with those obtained by Kim *et al.* (2006), who carried out the more thorough study encompassing diverse worldwide populations, either for *TAS1R3* or, as mentioned in the previous point, for *TAS1R1*, good agreement exists between data from the 2 studies. Also, as before referred, a major drawback of the study of Kim *et al.* (2006) was the very small sizes of the population samples analysed. Even so, similarly to the findings of this work, the authors observed that GGC was overwhelmingly present in their set of 8 populations, which included samples from Native Americans, Africans and Eurasians, being even the unique haplotype detected in 6 out the 8 populations, likely because they were quite undersized (from 8 to 20 individuals). Besides GGC, Kim *et al.* (2006) only detected other haplotypes in the Japanese and Cameroonians. Among the Japanese, GAC, which was the second most frequent haplotype detected in our study, occurred at 10% frequency; among the Cameroonians the additional haplotypes were GGT (25%) – also detected in this study, and AGC (2.5%) – which seems to be a rare haplotype not having been observed in the populations here examined.

A network was constructed with all *TAS1R3* haplotypes up to now detected, being shown in figure 15.

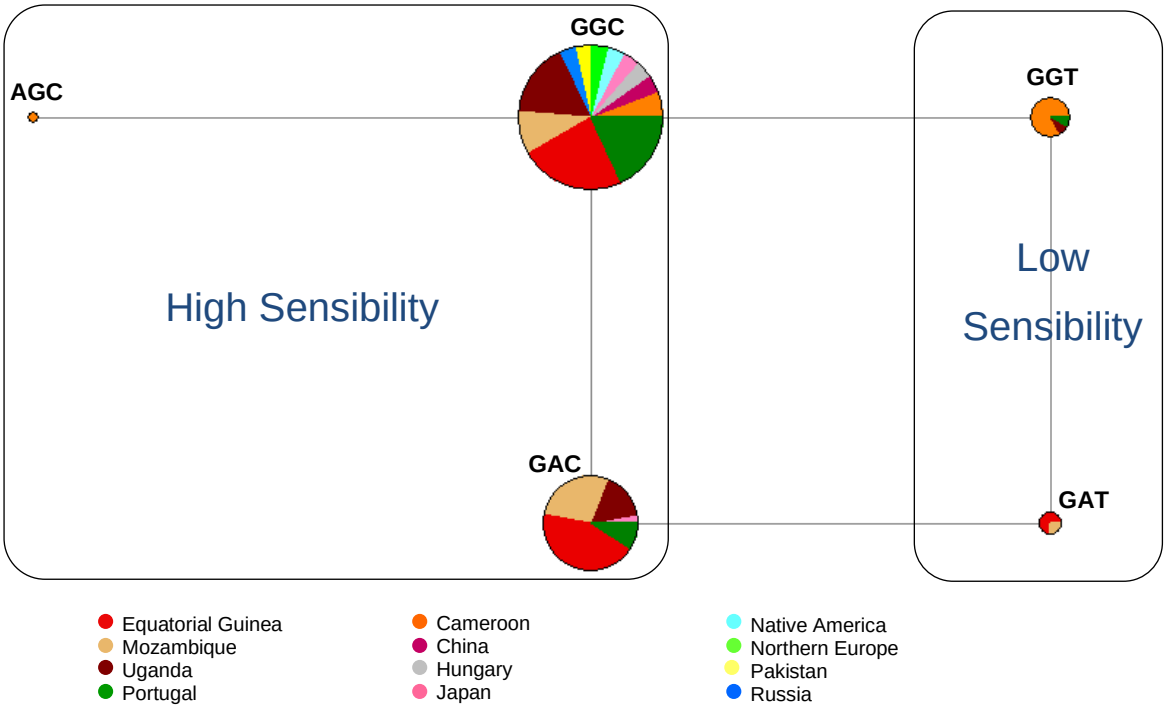


Fig. 15 – Network with *TAS1R3* haplotypes influencing the *umami* taste.

Overall, haplotypes conferring high sensibility to umami clearly prevail over those associated with decreased high sensibility. The presence of a reticulation in the network suggests that intragenic recombination likely accounted for the generation of the GAT haplotype combination.

Concerning the distribution of the combined frequency of high (GGC+GAC+AGC) and low sensibility haplotypes (GGT+GAT), in non-African populations low sensibility haplotypes were only found in the Portuguese studied in this work, among whom GGT was present at 6.1% frequency.

Within African populations, no major differences in the distribution of high and low sensibility haplotypes were observed, excepting the findings for the Cameroonian sample. High sensibility haplotypes represented 98% of chromosomes in Equatorial Guinea, 98.5% in Mozambique and attained the highest value in Uganda, 99%. Yet, in Cameroon their frequency was the lowest up to now register in any human population, 75%, although the value must be carefully considered, due to the small size of this sample (20 individuals).

So, in the African context, the genetic variation here assessed at *TAS1R3*, which is thought to influence *umami* taste perception, does not appears to be correlated with lifestyle or even geographical location of populations.

This conclusion can also be drawn from the analysis of pairwise F_{ST} genetics distances between populations, comprehending samples now studied plus those reported by Kim *et al.* (2006), which are presented in table 19.

Table 19 – F_{ST} values correspondent to *TAS1R3* haplotypes. Mz: Mozambique; EG: Ca: Cameroonian; EG: Equatorial Guinea; UG: Uganda; Ch: Chinese; NA: Native American; Hu: Hungarian; Pa: Pakistani; Ja: Japanese; Ru: Russian; PT: Portugal; NE: Northern European.

	MZ	Ca	EG	UG	Ch	NA	Hu	Pa	Ja	Ru	PT	NE
MZ	*											
Ca	**0.1560**	*										
EG	0.0347	**0.1429**	*									
UG	0.0978	**0.1558**	0.0443	*								
Ch	0.2046	0.2000	0.1282	0.0755	*							
NA	0.2046	0.2000	0.1282	0.0755	0.0285	*						
Hu	0.2046	0.2000	0.1282	0.0755	0.0285	0.0285	*					
Pa	0.1924	0.1839	0.1204	0.0671	0.0285	0.0285	0.0285	*				
Ja	0.0857	0.1296	0.0356	0.0000	0.0811	0.0811	0.0811	0.0653	*			
Ru	0.2046	0.2000	0.1282	0.0755	0.0285	0.0285	0.0285	0.0285	0.0811	*		
PT	**0.2182**	**0.1674**	**0.1236**	0.0642	0.0379	0.0379	0.0379	0.0379	0.0303	0.0379	*	
NE	0.2046	0.2000	0.1282	0.0755	0.0285	0.0285	0.0285	0.0285	0.0285	0.0285	0.0379	*

The significant values after Bonferroni's correction are in blue.

The stronger genetic differentiations were always registered in pairs involving an African and a non-African population. After applying the correction for multiple tests, few results maintained the significance, among which were the distances between Portugal and Mozambique, Cameroon and Equatorial Guinea. However, these significant genetic distances ranged considerably, from 21.8% (Portugal-Mozambique) to its half, 12.4% (Portugal – Equatorial Guinea). Within Africans, significant differentiations were only observed between pairs involving the Cameroonian sample, illustrating the singular genetic profile of this population (average genetic distance with other African populations of 15.2%).

Concerning AMOVA, the most relevant results are presented in table 20. Despite the several criteria used for grouping populations, proportions of variation among groups were only significant when the continent of origin was considered or when Africans were confronted with non-Africans. Using the latter criterion, the percentage of variation among groups was higher, significantly explaining 9.03% of the variation observed for *TAS1R3*.

Table 20 – AMOVA results correspondent to groups formed with 12 populations' data from different regions and ethnics.

Source of Variation	Percentage of variation	*P*-value
Geography (by continent)		
Among groups	6.59	0.0234± 0.0013
Among populations within groups	4.60	0.0002± 0.0002
Within populations	88.82	0.0000±0.0000
Africa vs. Non-Africa		
Among groups	9.03	0.0018±0.0004
Among populations within groups	3.66	0.0003±0.0002
Within populations	87.31	0.0000±0.0000

According to a recent study, *umami* taste perception seems to be linked with obesity in humans (Donaldson *et al.,* 2009). Being so, it would be important to obtain a better understanding of the genetic factors underlying *umami* taste sensibility, as well as of the selective forces, if they really were exerted, contributing to the current patterns of genetic variations implied in the ability to sense the *umami* taste. This characteristic seems, therefore, to be an important piece to the health-care debate.

1.3. Sweet taste

Two SNPs were selected in the promoter of *TAS1R3*, a gene encoding for a subunit that together with the subunit coded by *TAS1R2* also forms a human sweet receptor heterodimer. The two SNPs were rs307355 (-C1572T) and rs35744813 (-C1266T), and they were chosen because a previous study provided evidence of being strongly correlated with the ability of people to correctly sort ascending concentrations of sucrose (Fushan *et al.*, 2009). Both are C/T variations, and individuals carrying the T alleles were proven to display reduced sensitivity to sucrose compared to those who carry C alleles at the two nucleotide positions (Fushan *et al.*, 2009). For simplicity, hereafter, the CC haplotype will be referred to as the sucrose taster, the TT as the non-taster, whereas other arrangements will be considered intermediate haplotypes.

In table 21 are presented the haplotype frequencies estimated in the surveyed populations.

Table 21 – *TAS1R3* haplotype frequencies and diversities obtained for sweet taste.

Sample / Haplotype	Uganda (n=48)	Angola (n=17)	Mozambique (n=34)	Equatorial Guinea (n=76)	Portugal (n=49)
CC	0.2708±0.0472	0.0882±0.0492	0.3088±0.0558	0.3668±0.0407	0.7959±0.0403
CT	0.1667±0.0390	0.1471±0.0609	0.2059±0.0493	0.1661±0.0306	0.0510±0.0220
TC	-	-	-	0.0148±0.0405	-
TT	0.5625±0.0517	0.7647±0.0730	0.4853±0.0600	0.4523±0.0108	0.1531±0.0356
Haplotype Diversity	0.5885±0.0342	0.3999±0.0711	0.6361±0.0294	0.6352±0.0181	0.3440±0.0547

A remarkable difference was observed between Portugal and any of the African populations: whereas the taster CC haplotype clearly predominates in Portugal (79.6%), the most frequent one in Africans was always the non-taster TT (varying from 45.2% to 76.5%). Besides these two haplotypes, the intermediate CT was as well shared by all populations, occurring at low frequency in Portugal (5.1%) while reaching moderate values in Africans (14.7%-20.6%). In Equatorial Guinea, the rare intermediate TC haplotype was also detected at 1.5% frequency.

These data are entirely consistent with those reported by Fushan *et al.* (2009), who performed a worldwide survey and similarly detected strong discrepancy in the frequencies of the CC and TT haplotypes between Europeans and Africans. Fushan *et al.* (2009) further studied an Asian sample, which showed much tighter affinities with the European sample than with the African one. CT haplotype was not found in Asians and its frequency was very low (0.5%) in the European individuals in comparison to the

Portuguese population, where this intermediate haplotype was much higher frequency (5.1%).

Figure 16 shows the network of haplotypes and their frequencies in the populations studied in this work and in those screened by Fushan *et al*. (2009). The network shows up that at a broad geographical level there are two major haplotypes, which interestingly are the opposite CC (taster) and TT (non-taster). Although both are distributed across all geographical regions, CC is clearly overrepresented in Europe and Asia, while TT is much more frequent in Africa. The intermediate CT haplotype is rarely found out of Africa, occurring mainly in African populations, usually at moderate frequencies. The other intermediate haplotype, TC, here detected for the first time, was uniquely observed in Equatorial Guinea. Probably this haplotype combination has arisen by intragenic recombination between chromosomes harbouring the CC and TT haplotypes.

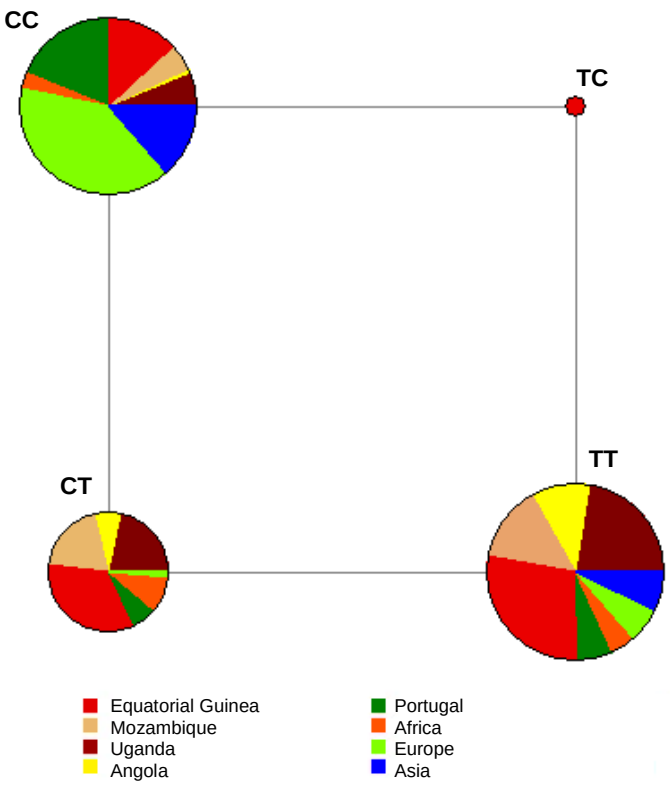


Fig. 16 – Network with *TAS1R3* haplotypes correspondent to sweet taste.

The F_{ST} distances between pair of populations studied in this work and in that of Fushan *et al.* (2009) are presented in table 22. All distances between pairs involving an African and a non-African population were highly significant, even under the correction for multiple tests. The range of genetic distances obtained, 19%-72% (average value 39%) reflects well the strong differentiation between African and non-African

populations for the two tested variations at *TAS1R3* that influence sensitivity to sweet compounds. This differentiation is clearly visualized in the MDS plot of the F_{ST} distances, presented in figure 17, where two well defined clusters can be seen, one grouping only African populations and the other the Eurasians ones.

Table 22 – F_{ST} values correspondent to *TAS1R3* haplotypes involved in sweet perception.

	Mozambique	Uganda	Angola	Equatorial Guinea	Africa	Europe	Asia	Portugal
Mozambique	*							
Uganda	0.0042	*						
Angola	**0.1039**	0.0647	*					
Equatorial Guinea	0.0039	0.0175	**0.1319**	*				
Africa	0.0033	0.0350	**0.2011**	0.0000	*			
Europe	**0.4930**	**0.5154**	**0.7277**	**0.3794**	**0.4520**	*		
Asia	**0.2620**	**0.2983**	**0.5281**	**0.1946**	**0.1985**	**0.0694**	*	
Portugal	**0.2893**	**0.3303**	**0.5453**	**0.2198**	**0.2151**	0.0417	0.0074	*

The significant values after Bonferroni's correction are in blue.

Concentrating now in the African populations, genetic distances between them are typically small and accordingly non-significant, excepting those involving Angola (table 22). Angola presents an unusual high frequency of the non-taster TT haplotype (76.5%), differentiating it, often significantly, from the other African populations. The position of Angola in the MDS plot (figure 17) illustrates well the singularity of this population in the African context.

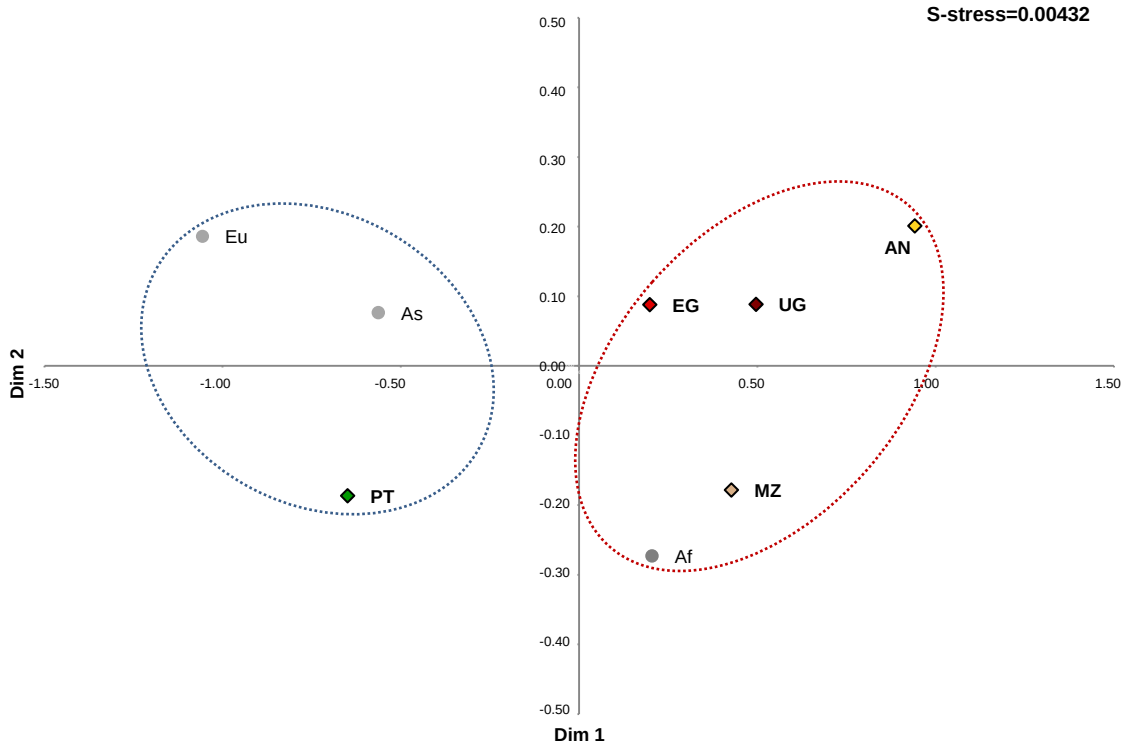


Fig. 17 – Multidimensional scaling plot of F_{ST} values to *TAS1R3* sweet haplotypes. AN: Angola; EG: Equatorial Guinea; MZ: Mozambique; UG: Uganda; PT: Portugal; Eu: European; As: Asian; Af: African.

Apart from Angola, the remaining African populations tend to be randomly positioned in their cluster, indicating that the distribution of the two variations that account for sweet taste perception does not seem related with lifestyle or even geography inside Africa.

The AMOVA results add support to this interpretation (table 23). In fact, the proportion of variation among groups was only significant when the groups were defined by continent of origin or when Africans were contrasted with non-Africans. Under the later criterion, the percentage of variation among groups achieved a more elevated value (34.29%), but in both cases differences among groups explained more than 30% of the total variation observed at the two SNPs. That proportion represented the highest differentiation among groups registered in the entire set of AMOVAs performed in this study concerning the different genetic variations investigated.

Table 23 – AMOVA results corresponding to groups formed with 8 populations' data from different regions and ethnics.

Source of Variation	Percentage of variation	*P*-value
Geography (by continent)		
Among groups	30.93	0.0047± 0.0007
Among populations within groups	2.80	0.0001± 0.0001
Within populations	66.27	0.0000±0.0000
Africa vs. Non-Africa		
Among groups	34.29	0.0169±0.0013
Among populations within groups	2.40	0.0000±0.0000
Within populations	63.31	0.0000±0.0000

The distribution of low and high sweet sensitivity haplotypes among human populations is suggestive that some kind of selective pressure might have been exerted favouring different genetic compositions in Africans and non-Africans. Fushan *et al*. (2009) hypothesised that such pressure could be related with the differential geographical distribution of sugar-rich aliments during human evolution. Assuming that initially carbohydrate-rich plants, like sugar cane and sugar beets, grew mainly in tropical latitudes, the authors admitted that the ability to taste sugars at lower concentrations might have been an important factor for human survival in cold geographical regions. Thus, high sweet sensitivity haplotypes could be favourable in the non-tropical geographical regions, driving their frequency to high values in those populations.

Whether selection acted or not in the genetic variations influencing sweet sensitivity is a question that still needs deep investigation. Yet, if the distribution of high and low sweet sensitivity haplotypes indeed represents genetic adaptations, no evidence was here obtained that such adaptations could be related with lifestyle of populations.

2. The role of lifestyle and geography in the different *loci*

One of the aims of this study was to evaluate whether some genetic variations that influence taste perception could be somehow related with diet/lifestyle of populations. A major difficulty to perform that investigation was the absence of data available worldwide for populations with different lifestyles, hindering so a deep analysis of the issue. Actually, assessment of the role of lifestyle in global patterns of genetic diversity was only possible to conduct for the two genes implicated in bitter taste perception *TAS2R16 and TAS2R38*. However, only for *TAS2R16* an indication appeared that its diversity could be somehow influenced by mode of subsistence. In fact, when AMOVA was performed splitting African populations according to their lifestyle, a significant value of variation among groups was observed, although the proportion of variation explainable by lifestyle was very small, around 3%. Furthermore, this proportion diminished, and lost its significance, when the analysis was extended to populations from outside Africa with similar lifestyles.

So, if lifestyle played any role in shaping diversity at this *locus*, the influence was not important enough to leave signs able to be clearly captured by our study's strategy.

For this *locus*, an interesting relationship was observed in Africa between the frequency distribution of the studied SNP and the Bantu expansions through sub-Saharan regions. Since Bantu people comprise the majority of African agriculturalists, this parallelism indicates that demographic processes must have had a major contribution to the current day pattern of distribution of the tested variation.

Concerning the remaining three genes, no minor support was detected that diet/lifestyle could be a variable determining allele frequencies, at least in Africa, since the lack of data for non-African populations with distinct lifestyles hampered to draw broader conclusions.

Contrarily, geography was found to be significantly correlated with the distribution of diversity in all *loci* analysed, although the relationship varied from *locus* to *locus*, as it was evidenced by the AMOVA results.

At a worldwide level, the differential role of geography in modelling diversity at each gene, becomes quite clear looking for the results shown in table 24 provided by AMOVA when it was executed using the continent of origin (Africa, America, Asia, Europe) as the clustering system.

Table 24 – AMOVA results to all genes in study regarding continental groups.

Source of Variation	Percentage of variation (%)
***TAS2R16* (bitter)**	
Among groups	**18.09**
Among populations within groups	**3.32**
Within populations	**78.58**
***TAS2R38* (bitter)**	
Among groups	**5.19**
Among populations within groups	**0.89**
Within populations	**93.92**
***TAS1R1* (*umami*)**	
Among groups	-1.31
Among populations within groups	**5.43**
Within populations	**95.88**
***TAS1R3* (*umami*)**	
Among groups	**6.59**
Among populations within groups	**4.60**
Within populations	**88.82**
***TAS1R3* (sweet)**	
Among groups	**30.93**
Among populations within groups	**2.80**
Within populations	**66.27**

The significant values are in blue.

Except for *TAS1R1*, in the remaining cases the amount of variation among groups was statistically significant, meaning that the continent of origin explains from 5.19% to 30.93% of those *loci* variation. For *TAS1R1*, the continent where populations are located does not have any impact on the pattern of genetic diversity.

Taking into account the entire set of variations, the average value of the proportion of variation among groups, is 11.88%. Interestingly, this average value is not far from the 13.2% proportion of differences among continents reported by Jorde *et al.* (2000) based upon the analysis of a wide number of SNPs. Since in the study of Jorde *et al.* (2000) most of the SNPs was supposedly neutral, the value reported by the author can be considered as a referential for the expected value of differentiation among

continents essentially explainable by the history of the dispersion of human populations throughout the world, which was rich in demographic events leading to successive bottlenecks and fluctuations in populations' size that strongly determined the current human genetic landscape (Jorde *et al.*, 2000; Handley *et al.*, 2007; Barbujani and Colonna, 2010).

The stronger departures from the average level of differences among continents were registered at the SNPs from *TAS1R1*, implied in the *umami* perception, and at those from *TAS1R3* associated with sweet perception (figure 18).

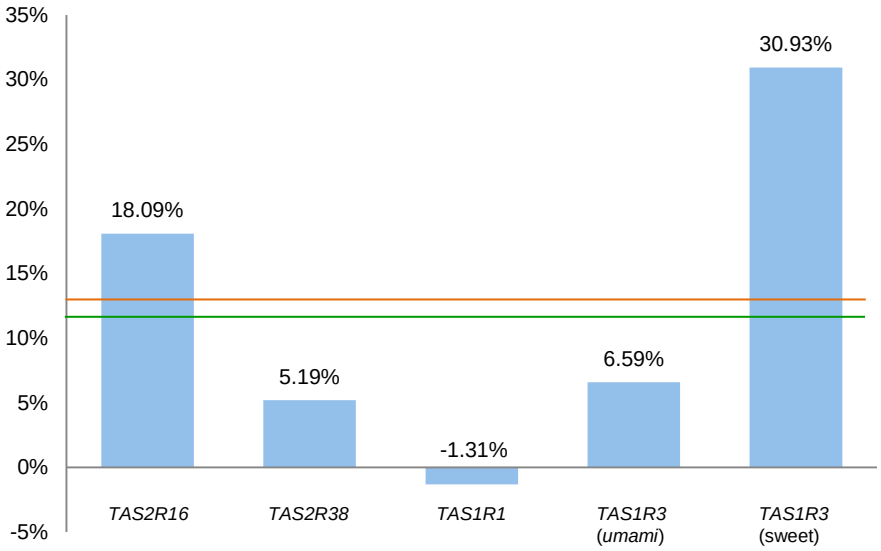


Fig. 18 - Graphic representation of the percentage of variation among groups to all genes studied. The green line represents the average value of variation obtained in our study (11.88%) and the orange the percentage of variation reported by Jorde *et al.* (2000) (13.20%).

However, in these two cases the deviations were opposite. Whereas *TAS1R1* is characterised by and unusual homogeneity across continents, the sweet-related variations at *TAS1R3* show levels of differentiation among the major world regions almost three times higher than commonly observed.

These findings deserve obvious attention. Although being impossible for now to exclude that the two strongest departures are only extreme cases that by chance occurred along the history of human populations, the departures are indeed high enough to support previous indications that the variations at *TAS1R1* and the sweet-related ones at *TAS1R3*, had been under selective pressures (Kim *et al.*, 2006; Fushan *et al.*, 2009).

If such selective pressures were indeed exerted, for the *umami*-related SNPs at *TAS1R1*, they do not seem to be environmental-related. In fact, the homogeneity

across the world suggests that some selective force, not dependent on geography, has acted maintaining at very high frequencies a haplotype associated to intermediate sensibility to *umami* perception (table 24). By the contrary, any possible selective factor affecting the sweet-related SNPs at *TAS1R3*, was likely related with variable(s) dependent on geography, determining the strong differentiation nowadays observed among populations from different continents, and especially among African and non-African regions (figure 17).

V. CONCLUSIONS

The present work represented a contribution to enlarge the knowledge about the genetic diversity in Africa that influences taste perception and its relationship with lifestyle, and associated diets, of populations.

The samples analysed included representatives of two modes of subsistence: a sample of pastoralists from Uganda and three samples of agriculturalists from Angola, Mozambique and Equatorial Guinea. A population sample from Portugal was also studied to act as a control of non-African origin, in order to facilitate a more clear interpretation of the data.

Till now, there is only one work published by Campbell and co-workers (2012), focusing the analysis of the correlation between different lifestyles practised in Africa and a taste related gene, which was *TAS2R38*, a gene that influences sensitivity to bitter compounds.

In this study, besides *TAS2R38*, three additional genes were examined in order to obtain a better picture of the genetic architecture underlying taste perception across African populations, as well as to obtain insights on how typical diets in populations were related with their genetic variability in taste perception.

Different SNPs previously implied in the phenotypic variability in *umami*, sweet and bitter taste perception were analysed, located in the following four genes: *TAS1R1*, *TAS1R3*, *TAS2R16* and *TAS2R38.* For their genotyping two PCR-multiplexes systems were successfully optimized.

One important achievement of this work was the characterization of five populations not studied before for the selected genetic variations. The data obtained constitutes an enrichment of the raw information essential to develop further studies, enrichment that is especially important regarding African populations, since they are clearly understudied up to date.

With the exception of *TAS1R1*, for the remaining tested *loci* African populations had higher genetic diversities compared to other populations. This was in conformity with previous works performed at worldwide scale, which clearly demonstrated that Africa is the continent harbouring the greatest level of genetic diversity (Campbell and Tishkoff, 2008).

At a worldwide level, it was only possible to analyse the influence of the mode of subsistence and genetic diversity for the two bitter genes here studied: *TAS2R16* and *TAS2R38.* However, for both genes we did not found any significant association between lifestyle and the corresponding patterns of worldwide allele distribution. For *TAS2R16*, a marginally significant association was detected when focusing the analysis in African populations, but yet lifestyle accounted no more than 3% to explain the variation within Africa.

A factor that highly influences the distribution patterns of the majority of the tested genetic variations is geography. Indeed, excluding the SNPs at *TAS1R1,* for which high homogeneity was observed across populations from different world regions*,* concerning the remaining variations the continent of origin was found to be correlated with the level of differentiation between populations.

However, some hints were also obtained indicating that geography and history of populations do not appear enough to explain the pattern of diversity at some of the variations. The worldwide distributions of variations at *TAS2R38, TAS1R1 and TAS1R3* raise the question of whether selection has not contributed to model diversity at those *loci*, although the nature of the hypothetical selective pressures remains completely obscure.

The utility of these polymorphisms, in cases of DNA phenotyping, must be more deeply explored in the future.

So, many further studies are needed to continue the line of investigation initiated with this work.

Future Directions

In order to reach another level of knowledge and information about the influence of population' lifestyles and genetic diversity at taste related variations, samples from African hunter-gatherers should be genotyped with the two multiplexes optimized in this work.

It also would be important to extend the genetic characterizations to populations from other continents that still maintaining distinct modes of subsistence.

Another study that in the future needs to be importantly performed, is to deeply investigate the relationship between genetic variation and phenotypic variability for the different categories of taste perception, a study that until now has been almost exclusively performed in European populations or in those of European descent.

Finally, besides testing only specific SNPs at each *locus*, it would be important to finely enlarge the genetic screening to the entire genomic sequences of the genes under study, in order to look for other signs of selection retained in those genes.

VI. BIBLIOGRAPHY

Bachmanov, A.A., Reed, D.R., Li, X., and Beauchamp, G.K. (2002). Genetics of sweet taste preferences. Pure and Applied Chemistry 74, 1135-1140.

Bachmanov, A.A., and Beauchamp, G.K. (2007). Taste receptor genes. Annual Review of Nutrition. pp 389-414.

Bachmanov, A.A., Bosak, N.P., Floriano, W.B., Inoue, M., Li, X., Lin, C., Murovets, V.O., Reed, D.R., Zolotarev, V.A., and Beauchamp, G.K. (2011). Genetics of sweet taste preferences. Flavour and Fragrance Journal 26, 286-294.

Balaresque, P.L., Ballereau, S.J., and Jobling, M.A. (2007). Challenges in human genetic diversity: demographic history and adaptation. Human Molecular Genetics 16, R134-R139.

Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16, 37-48.

Barbujani, G., and Colonna, V. (2010). Human genome diversity: frequently asked questions. Trends in Genetics 26, 285-295.

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., van der Veen, L., Hombert, J.-M., Quintana-Murci, L., and Comas, D. (2009). Genetic and Demographic Implications of the Bantu Expansion: Insights from Human Paternal Lineages. Molecular Biology and Evolution 26, 1581-1589.

Boughter, J.D., Jr., and Bachmanov, A.A. (2007). Behavioral genetics and taste. Bmc Neuroscience 8.

Bower, J. (1991). The Pastoral Neolithic of East-Africa. Journal of World Prehistory 5, 49-82.

Branicki, W., Liu, F., van Duijn, K., Draus-Barini, J., Pospiech, E., Walsh, S., Kupiec, T., Wojas-Pelc, A., and Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. Human Genetics 129, 443-454.

Bufe, B., Hofmann, T., Krautwurst, D., Raguse, J.D., and Meyerhof, W. (2002). The human TAS2R16 receptor mediates bitter taste in response to beta-glucopyranosides. Nature Genetics 32, 397-401.

Bufe, B., Breslin, P.A.S., Kuhn, C., Reed, D.R., Tharp, C.D., Slack, J.P., Kim, U.K., Drayna, D., and Meyerhof, W. (2005). The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. Current Biology 15, 322-327.

Butler, J.M. (2005). Forensic DNA typing: biology, technology, and genetics of STR markers. (Academic Press).

Campbell, M.C., and Tishkoff, S.A. (2008). African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. In Annual Review of Genomics and Human Genetics. pp 403-433.

Campbell, M.C., and Tishkoff, S.A. (2010). The Evolution of Human Genetic and Phenotypic Variation in Africa. Current Biology 20, R166-R173.

Campbell, M.C., Ranciaro, A., Froment, A., Hirbo, J., Omar, S., Bodo, J.-M., Nyambo, T., Lema, G., Zinshteyn, D., Drayna, D., et al. (2012). Evolution of Functionally Diverse Alleles Associated with PTC Bitter Taste Sensitivity in Africa. Molecular Biology and Evolution 29, 1141-1153.

Cavalli-Sforza, L.L. (1997). Genes, peoples, and languages. Proceedings of the National Academy of Sciences of the United States of America 94, 7719-7724.

Cavalli-Sforza, L.L., and Feldman, M.W. (2003). The application of molecular genetic approaches to the study of human evolution. Nature Genetics 33, 266-275.

Chandrashekar, J., Hoon, M.A., Ryba, N.J.P., and Zuker, C.S. (2006). The receptors and cells for mammalian taste. Nature 444, 288-294.

Chaudhari, N., Pereira, E., and Roper, S.D. (2009). Taste receptors for *umami*: the case for multiple receptors. American Journal of Clinical Nutrition 90, 738S-742S.

Chen, F.C., and Li, W.H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. American Journal of Human Genetics 68, 444-456.

Chen, Q.-Y., Alarcon, S., Tharp, A., Ahmed, O.M., Estrella, N.L., Greene, T.A., Rucker, J., and Breslin, P.A.S. (2009). Perceptual variation in *umami* taste and polymorphisms in TAS1R taste receptor genes. American Journal of Clinical Nutrition 90, 770S-779S.

de Krom, M., Bauer, F., Collier, D., Adan, R.A.H., and la Fleur, S.E. (2009). Genetic Variation and Effects on Human Eating Behavior. In Annual Review of Nutrition. pp 283-304.

DeSimone, J.A., and Lyall, V. (2006). Taste receptors in the gastrointestinal tract III. Salty and sour taste: sensing of sodium and protons by the tongue. American Journal of Physiology-Gastrointestinal and Liver Physiology 291, G1005-G1010.

Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. Nature 418, 700-707.

Diamond, J., and Bellwood, P. (2003). Farmers and their languages: The first expansions. Science 300, 597-603.

Donaldson, L.F., Bennett, L., Baic, S., and Melichar, J.K. (2009). Taste and weight: is there a link? The American journal of clinical nutrition 90, 800S-803S.

Drayna, D. (2005). Human taste genetics. In Annual Review of Genomics and Human Genetics. pp 217-235.

Drewnowski, A., and Rock, C.L. (1995). The influence of genetic taste markers on food acceptance. American Journal of Clinical Nutrition 62, 506-511.

Drewnowski, A. (2002). Genetic markers, taste responses, and food preferences. In Chemistry of Taste: Mechanisms, Behaviors, and Mimics, P.P.D. Given, ed., pp 52-64.

Enserink, M. (2011). Can This DNA Sleuth Help Catch Criminals? Science 331, 838-840.

Excoffier, L., and Lischer, H.E.L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10, 564-567.

Feeney, E., O'Brien, S., Scannell, A., Markey, A., and Gibney, E.R. (2011). Genetic variation in taste perception: does it have a role in healthy eating? Proceedings of the Nutrition Society 70, 135-143.

Feero, W.G., Guttmacher, A.E., and Collins, F.S. (2010). Genomic Medicine: Genomic Medicine - An Updated Primer. New England Journal of Medicine 362, 2001-2011.

Fushan, A.A., Simons, C.T., Slack, J.P., Manichaikul, A., and Drayna, D. (2009). Allelic Polymorphism within the *TAS1R3* Promoter Is Associated with Human Taste Sensitivity to Sucrose. Current Biology 19, 1288-1293.

Garcia-Bailo, B., Toguri, C., Eny, K.M., and El-Sohemy, A. (2009). Genetic Variation in Taste and Its Influence on Food Selection. Omics 13, 69-80.

Gignoux, C.R., Henn, B.M., and Mountain, J.L. (2011). Rapid, global demographic expansions after the origins of agriculture. Proceedings of the National Academy of Sciences of the United States of America 108, 6044-6049.

Griffiths, A., Wessler, S., Lewontion, R., and Carroll, S. B (2008) Introduction to genetic analysis. WH Freeman and Company.

Gupta, A.K. (2004). Origin of agriculture and domestication of plants and animals linked to early Holocene climate amelioration. Current Science 87, 54-59.

Handley, L.J.L., Manica, A., Goudet, J., and Balloux, F. (2007). Going the distance: human population genetics in a clinal world. Trends in Genetics 23, 432-439.

Hinrichs, A.L., Wang, J.C., Bufe, B., Kwon, J.M., Budde, J., Allen, R., Bertelsen, S., Evans, W., Dick, D., and Rice, J. (2006). Functional Variant in a Bitter-Taste Receptor (*hTAS2R16*) Influences Risk of Alcohol Dependence. The American Journal of Human Genetics 78, 103-111.

Hladik, C.M., Pasquet, P., and Simmen, B. (2002). New perspectives on taste and primate evolution: The dichotomy in gustatory coding for perception of beneficent versus noxious substances as supported by correlations among human thresholds. American Journal of Physical Anthropology 117, 342-348.

Holden, C.J. (2006). The spread of Bantu languages, farming, and pastoralism in sub-equatorial Africa.

Hollox, E.J., Poulter, M., Zvarik, M., Ferak, V., Krause, A., Jenkins, T., Saha, N., Kozlov, A.I., and Swallow, D.M. (2001). Lactase haplotype diversity in the Old World. American Journal of Human Genetics 68, 160-172.

Jobling, M.A., Hurles, M.E., Tyler-Smith, C. (2004). Human Evolutionary Genetic- Origins, Peoples & Disease. Garland Science.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T., and Batzer, M.A. (2000). The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. American Journal of Human Genetics 66, 979-988.

Keller, K.L., Steinmann, L., Nurse, R.J., and Tepper, B.J. (2002). Genetic taste sensitivity to 6-n-propylthiouracil influences food preference and reported intake in preschool children. Appetite 38, 3-12.

Kim, U.K., Breslin, P.A.S., Reed, D., and Drayna, D. (2004). Genetics of human taste perception. Journal of Dental Research 83, 448-453.

Kim, U.K., Wooding, S., Riaz, N., Jorde, L.B., and Drayna, D. (2006). Variation in the human TAS1R taste receptor genes. Chemical Senses 31, 599-611.

Kurihara, K., and Kashiwayanagi, M. (2000). Physiological studies on umami taste. Journal of Nutrition 130, 931S-934S.

Kurihara, K. (2009). Glutamate: from discovery as a food flavor to role as a basic taste (*umami*). American Journal of Clinical Nutrition 90, 719S-722S.

Lareu, M.V., Phillips, C.P., Carracedo, A., Lincoln, P.J., Syndercombe Court, D., and Thomson, J.A. (1994). Investigation of the STR *locus* HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. Forensic Science International 66, 41-52.

Laugerette, F., Passilly-Degrace, P., Patris, B., Niot, I., Febbraio, M., Montmayeur, J.P., and Besnard, P. (2005). CD36 involvement in orosensory detection of dietary lipids, spontaneous fat preference, and digestive secretions. Journal of Clinical Investigation 115, 3177-3184.

Laugerette, F., Gaillard, D., Passilly-Degrace, P., Niot, I., and Besnard, P. (2007). Do we taste fat? Biochimie 89, 265-269.

Li, H., Pakstis, A.J., Kidd, J.R., and Kidd, K.K. (2011). Selection on the Human Bitter Taste Gene, *TAS2R16*, in Eurasian Populations. Human Biology 83, 363-377.

Li, X.D., Staszewski, L., Xu, H., Durick, K., Zoller, M., and Adler, E. (2002). Human receptors for sweet and *umami* taste. Proceedings of the National Academy of Sciences of the United States of America 99, 4692-4696.

Lindemann, B., Ogiwara, Y., and Ninomiya, Y. (2002). The discovery of *umami*. Chemical Senses 27, 843-844.

LopezJimenez, N.D., Cavenagh, M.M., Sainz, E., Cruz-Ithier, M.A., Battey, J.F., and Sullivan, S.L. (2006). Two members of the TRPP family of ion channels, Pkd1l3 and Pkd2l1, are co-expressed in a subset of taste receptor cells. Journal of Neurochemistry 98, 68-77.

Luca, F., Perry, G.H., and Di Rienzo, A. (2010). Evolutionary Adaptations to Dietary Changes. In Annual Review of Nutrition, Vol 30, R.J. Cousins, ed., pp 291-314.

Lugaz, O., Pillias, A.M., and Faurion, A. (2002). A new specific ageusia: Some humans cannot taste L-glutamate. Chemical Senses 27, 105-115.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1989). Molecular cloning : a laboratory manual / J. Sambrook, E.F. Fritsch, T. Maniatis. In. (New York: Cold Spring Harbor Laboratory Press.

Marshall, F. (1990). Origins of specialized pastoral production in East-Africa. American Anthropologist 92, 873-894.

Milton, K. (2003). The critical role played by animal source foods in human (Homo) evolution. Journal of Nutrition 133, 3886S-3892S.

Nelson, G., Chandrashekar, J., Hoon, M.A., Feng, L.X., Zhao, G., Ryba, N.J.P., and Zuker, C.S. (2002). An amino-acid taste receptor. Nature 416, 199-202.

Nielsen, R. (2005). Molecular signatures of natural selection. Annual Review of Genetics. pp 197-218.

Pakendorf, B., Bostoen, K., and de Filippo, C. (2011). Molecular perspectives on the Bantu expansion: a synthesis. Language Dynamics and Change 1, 50-88.

Patin, E., and Quintana-Murci, L. (2008). Demeter's legacy: rapid changes to our genome imposed by diet. Trends in Ecology & Evolution 23, 56-59.

Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., et al. (2007). Diet and the evolution of human amylase gene copy number variation. Nature Genetics 39, 1256-1260.

Powell, K.B., Mortensen, H., and Tishkoff, S.A. (2003). The evolution of lactase persistence in African populations. American Journal of Physical Anthropology, 170-170.

Rajeevan, H., Soundararajan, U., Kidd, J.R., Pakstis, A.J., and Kidd, K.K. (2012). ALFRED: an allele frequency resource for research and teaching. Nucleic acids research 40, D1010-D1015.

Raliou, M., Wiencis, A., Pillias, A.M., Planchais, A., Eloit, C., Boucher, Y., Trotier, D., Montmayeur, J.P., and Faurion, A. (2009). Nonsynonymous single nucleotide polymorphisms in human tas1r1, tas1r3, and mGluR1 and individual taste sensitivity to glutamate. The American journal of clinical nutrition 90, 789S-799S.

Reed, F.A., and Tishkoff, S.A. (2006). African human diversity, origins and migrations. Current Opinion in Genetics & Development 16, 597-605.

Roper, S.D. (2007). Signal transduction and information processing in mammalian taste buds. Pflugers Archiv-European Journal of Physiology 454, 759-776.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132, 365-386

Sherratt, A., and Sherratt, S. (1988). The archaeology of Indo-European: an alternative view. Antiquity 62, 584-595.

Shigemura, N., Shirosaki, S., Sanematsu, K., Ogiwara, Y., Kawai, M., Yoshida, R., and Ninomiya, Y. (2009a). Functional analysis of T1R1/T1R3 *umami* taste receptor variants. Neuroscience Research 65, S176-S176.

Shigemura, N., Shirosaki, S., Sanematsu, K., Yoshida, R., and Ninomiya, Y. (2009b). Genetic and Molecular Basis of Individual Differences in Human *Umami* Taste Perception. Plos One 4.

Soranzo, N., Bufe, B., Sabeti, P.C., Wilson, J.F., Weale, M.E., Marguerie, R., Meyerhof, W., and Goldstein, D.B. (2005). Positive selection on a high-sensitivity allele of the human bitter-taste receptor *TAS2R16*. Current Biology 15, 1257-1265.

Tepper, B.J. (2008). Nutritional implications of genetic taste variation: The role of PROP sensitivity and other taste phenotypes. In Annual Review of Nutrition. pp 367-388.

Tepper, B.J., White, E.A., Koelliker, Y., Lanzara, C., d'Adamo, P., and Gasparini, P. (2009). Genetic Variation in Taste Sensitivity to 6-n-Propylthiouracil and Its Relationship to Taste Perception and Food Selection. In International Symposium on Olfaction and Taste, T.E. Finger, ed., pp 126-139.

The International HapMap Consortium (2010). Integrating common and rare genetic variation in diverse human populations.Nature 467, 52-58.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.-M., Doumbo, O., Ibrahim M., Juma A.T., Kotze M.J., Lema G., Moore J. H., Mortensen H., Nyambo T. B., Omar S. A., Powell K., Pretorius G. S., Smith M. W., Thera M. A., Wambebe C., Weber J. L., Williams S. M. (2009). The Genetic Structure and History of Africans and African Americans. Science 324, 1035-1044.

Tornwall, O., Silventoinen, K., Keskitalo-Vuokko, K., Perola, M., Kaprio, J., and Tuorila, H. (2012). Genetic contribution to sour taste preference. Appetite 58, 687-694.

Walsh, S., Liu, F., Ballantyne, K.N., van Oven, M., Lao, O., and Kayser, M. (2011a). IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Science International: Genetics 5, 170-180.

Walsh, S., Lindenbergh, A., Zuniga, S.B., Sijen, T., de Knijff, P., Kayser, M., and Ballantyne, K.N. (2011b). Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Science International-Genetics 5, 464-471.

Wise, P.M., Hansen, J.L., Reed, D.R., and Breslin, P.A.S. (2007). Twin study of the heritability of recognition thresholds for sour and salty taste. Chemical Senses 32, 749-754.

Wooding, S., Kim, U.K., Bamshad, M.J., Larsen, J., Jorde, L.B., and Drayna, D. (2004). Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. American Journal of Human Genetics 74, 637-646.

Wooding, S. (2005). Evolution: A study in bad taste? Current Biology 15, R805-R807.

Wooding, S. (2006). Phenylthiocarbamide: A 75-year adventure in genetics and natural selection. Genetics 172, 2015-2023.

Wooding, S., Gunn, H., Ramos, P., Thalmann, S., Xing, C., and Meyerhof, W. (2010). Genetics and Bitter Taste Responses to Goitrin, a Plant Toxin Found in Vegetables. Chemical Senses 35, 685-+.

Wooding, S. (2011). Signatures of Natural Selection in a Primate Bitter Taste Receptor. Journal of Molecular Evolution 73, 257-265.

Ye, K., and Gu, Z. (2011). Recent advances in understanding the role of nutrition in human genome evolution. Advances in nutrition (Bethesda, Md) 2, 486-496.

Zhao, G.Q., Zhang, Y.F., Hoon, M.A., Chandrashekar, J., Erlenbach, I., Ryba, N.J.P., and Zuker, C.S. (2003). The receptors for mammalian sweet and umami taste. Cell 115, 255-266.

VII. APPENDIX

Annex 1. Supplementary data for the eleven *loci* studied.

Table S1 – *Locus* by *locus* Heterozygosity and correspondent *P*-values for all populations studied.

	Uganda			Angola			Mozambique			Equatorial Guinea			Portugal		
Locus	**Ho**	**He**	***P*-value**	**Ho**	**He**	***P*-value**	**Ho**	**He**	***P*-value**	**Ho**	**He**	***P*-value**	**Ho**	**He**	***P*-value**
C329T	monomorphic *locus*			monomorphic *locus*			monomorphic *locus*			0.0395	0.0390	1.0000±0.0000	0.0408	0.0404	1.0000±0.0000
G1114A	0.1042	0.0998	1.0000±0.0000	monomorphic *locus*			0.0303	0.0303	1.0000±0.0000	0.1184	0.1122	1.0000±0.0000	0.2245	0.2014	1.0000±0.0000
G13A	0.0204	0.0204	1.0000±0.0000	monomorphic *locus*			monomorphic *locus*			monomorphic *locus*			monomorphic *locus*		
G740A	0.1429	0.2014	0.0920±0.0003	0.3182	0.4598	0.1781±0.0004	0.4118	0.4214	1.0000±0.0000	0.3290	0.3422	0.7410±0.0004	0.0204	0.0204	1.0000±0.0000
C2269T	0.0204	0.0204	1.0000±0.0000	monomorphic *locus*			0.0294	0.0865	0.0147±0.0001	0.0395	0.0390	1.0000±0.0000	0.1225	0.1161	1.0000±0.0000
-C1572T	0.3674	0.4949	0.0837±0.0003	0.2353	0.3708	0.1769±0.0004	0.4412	0.5070	0.5044±0.0005	0.4079	0.5011	0.1120±0.0003	0.2245	0.2619	0.2932±0.0005
-C1266T	0.2857	0.3939	0.0692±0.0003	0.0588	0.1658	0.0910±0.0003	0.4412	0.4333	1.0000±0.0000	0.3158	0.4751	0.0066±0.0001	0.2449	0.3282	0.0889±0.0003
G516T	0.2500	0.2518	1.0000±0.0000	0.4286	0.4936	0.6596±0.0005	0.3438	0.4241	0.3950±0.0005	0.3158	0.3904	0.1346±0.0003	monomorphic *locus*		
C145G	0.6250	0.5018	0.1418±0.0004	0.3478	0.5101	0.2073±0.0004	0.5294	0.5057	1.0000±0.0000	0.5658	0.5025	0.3571±0.0005	0.5306	0.5033	0.7767±0.0004
C785T	0.4565	0.4193	0.7241±0.0005	0.2000	0.4308	0.0263±0.0002	0.5484	0.4321	0.2077±0.0004	0.4079	0.4196	0.7898±0.0004	0.5306	0.4949	0.7706±0.0004
G886A	0.6458	0.5156	0.0825±0.0002	0.4286	0.5122	0.6628±0.0005	0.5938	0.5035	0.4740±0.0005	0.5526	0.5019	0.4887±0.0005	0.5306	0.4999	0.7747±0.0004

P=0.0045 (after the Bonferroni correction for multiple tests)

Table S2 – Genotypic data for all populations studied.

I. Angola

	TAS1R1		***TAS1R3***					***TAS2R16***	***TAS2R38***		
	C329T	**G1114A**	**G13A**	**G740A**	**C2269T**	**-C1572T**	**-C1266T**	**G516T**	**C145G**	**C785T**	**G886A**
A1	n/a	n/a	n/a	G	n/a	n/a	n/a	n/a	G	n/a	n/a
	n/a	n/a	n/a	G	n/a	n/a	n/a	n/a	C	n/a	n/a
A2	C	n/a	n/a	G	C	n/a	n/a	n/a	C	n/a	n/a
	C	n/a	n/a	G	C	n/a	n/a	n/a	C	n/a	n/a
A3	C	G	G	n/a	C	T	T	T	G	T	G
	C	G	G	n/a	C	T	T	T	C	C	A
A4	n/a	n/a	n/a	G	C	n/a	n/a	n/a	G	n/a	n/a
	n/a	n/a	n/a	G	C	n/a	n/a	n/a	G	n/a	n/a
A5	C	n/a	G	G	C	T	T	T	C	T	G
	C	n/a	G	A	C	T	T	G	C	T	A
A6	C	n/a	n/a	n/a	C	T	T	T	C	n/a	G
	C	n/a	n/a	n/a	C	T	T	G	C	n/a	G
A7	C	n/a	n/a	A	C	n/a	n/a	T	G	n/a	n/a
	C	n/a	n/a	A	C	n/a	n/a	T	G	n/a	n/a
A8	n/a	n/a	G	G	C	T	T	T	G	T	A
	n/a	n/a	G	A	C	T	T	G	G	T	A
A9	C	n/a	n/a	G	C	T	T	T	C	T	G
	C	n/a	n/a	G	C	T	T	G	C	C	A
A10	C	n/a	n/a	A	C	T	T	T	G	C	A
	C	n/a	n/a	A	C	C	T	G	G	C	A
A11	C	n/a	G	G	C	T	T	T	G	C	G
	C	n/a	G	G	C	T	T	T	C	C	G
A12	C	n/a	n/a	A	n/a	n/a	n/a	G	G	C	A
	C	n/a	n/a	A	n/a	n/a	n/a	G	C	C	A
A13	n/a	n/a	G	G	C	T	T	n/a	G	T	A
	n/a	n/a	G	G	C	T	T	n/a	G	T	A
A14	C	n/a	n/a	n/a	n/a	n/a	n/a	T	n/a	C	G
	C	n/a	n/a	n/a	n/a	n/a	n/a	T	n/a	C	A
A15	C	n/a	G	G	C	T	T	n/a	G	n/a	n/a
	C	n/a	G	A	C	C	T	n/a	C	n/a	n/a
A16	C	n/a	n/a	G	C	T	T	T	G	C	G
	C	n/a	n/a	A	C	T	T	G	C	C	A

A17	C	n/a	n/a	n/a	n/a	n/a	n/a	G	n/a	T	A
	C	n/a	n/a	n/a	n/a	n/a	n/a	G	n/a	T	A
A18	C	n/a	n/a	G	n/a	n/a	n/a	G	C	C	G
	C	n/a	n/a	A	n/a	n/a	n/a	G	C	C	G
A19	C	n/a	n/a	G	C	n/a	n/a	T	n/a	C	G
	C	n/a	n/a	A	C	n/a	n/a	T	n/a	C	A
A20	n/a	n/a	n/a	G	C	C	C	G	G	C	G
	n/a	n/a	n/a	G	C	C	C	G	C	C	A
A21	C	n/a	G	G	C	T	T	T	G	T	A
	C	n/a	G	G	C	C	C	G	G	C	A
A22	C	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	C	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
A23	C	n/a	n/a	A	n/a	T	T	T	C	C	G
	C	n/a	n/a	A	n/a	T	T	T	C	C	G
A24	C	n/a	G	G	C	T	T	T	G	C	G
	C	n/a	G	G	C	T	T	T	G	C	A
A25	C	n/a	n/a	G	C	n/a	n/a	T	C	C	G
	C	n/a	n/a	A	C	n/a	n/a	T	C	C	G
A26	C	n/a	n/a	G	C	C	T	T	C	C	G
	C	n/a	n/a	G	C	C	T	G	C	C	G
A27	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	T	G	C	C	A

II. Mozambique

	TAS1R1		*TAS1R3*					*TAS2R16*	*TAS2R38*		
	C329T	**G1114A**	**G13A**	**G740A**	**C2269T**	**-C1572T**	**-C1266T**	**G516T**	**C145G**	**C785T**	**G886A**
M1	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	A	C	C	C	T	C	C	A
M2	C	G	G	G	C	T	T	T	C	C	G
	C	G	G	A	C	T	T	T	C	C	G
M3	C	G	G	G	C	C	C	G	G	C	G
	C	G	G	G	C	C	C	G	C	C	A
M4	C	G	G	G	T	T	T	T	G	C	G
	C	G	G	A	C	T	T	G	C	C	A
M5	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	C	A
M6	C	G	G	G	C	C	T	T	G	T	G
	C	G	G	G	C	C	C	G	C	C	A
M7	C	G	G	G	C	T	T	T	C	T	G
	C	G	G	G	C	C	T	G	C	C	G
M8	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	A	C	T	T	T	C	C	A
M9	C	G	G	A	C	T	T	T	G	T	A
	C	G	G	A	C	T	T	G	G	C	A
M10	C	G	G	G	C	T	T	G	C	T	G
	C	G	G	A	C	T	T	G	C	C	G
M11	C	G	G	A	C	T	T	T	G	C	G
	C	G	G	A	C	T	T	G	C	C	A
M12	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	A	C	C	T	T	C	C	A
M13	C	G	G	G	C	T	T	T	G	T	A
	C	G	G	A	C	C	C	T	G	T	A
M14	C	G	G	A	C	T	T	T	G	C	G
	C	G	G	A	C	T	T	G	G	C	A
M15	C	G	G	G	C	C	C	T	G	T	G
	C	G	G	A	C	C	C	T	C	C	A
M16	C	G	G	G	C	T	T	T	C	C	G
	C	A	G	G	C	C	C	G	C	C	G
M17	n/a	G	G	G	C	T	T	n/a	C	n/a	n/a
	n/a	G	G	A	C	C	C	n/a	C	n/a	n/a
M18	C	G	G	G	C	T	T	G	G	T	G
	C	G	G	G	C	C	C	G	C	C	A
M19	C	G	G	G	C	C	T	T	G	n/a	A
	C	G	G	G	C	C	C	T	G	n/a	A
M20	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	A	C	T	T	G	C	C	A
M21	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	A	C	C	T	T	C	C	A
M22	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	T	T	G	C	C	A
M23	C	G	G	G	C	T	T	T	C	C	G
	C	G	G	A	C	C	C	T	C	C	G
M24	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	C	C	G	C	C	A
M25	C	G	G	G	C	C	T	T	C	T	G
	C	G	G	G	C	C	C	T	C	C	G

M26	C	G	G	G	C	C	T	T	C	T	G
	C	G	G	G	C	C	C	T	C	C	G
M27	C	G	G	G	C	C	T	T	G	C	G
	C	G	G	G	C	C	T	T	C	C	A
M28	C	G	G	G	C	C	T	T	G	T	G
	C	G	G	G	C	C	C	T	C	C	A
M29	C	G	G	G	C	T	T	T	G	C	A
	C	G	G	A	C	C	T	T	G	C	A
M30	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	C	T	T	C	C	A
M31	C	G	G	G	G	T	T	T	C	C	G
	C	G	G	G	G	C	T	T	C	C	G
M32	C	G	G	G	C	C	T	T	G	T	G
	C	G	G	G	C	C	C	G	C	C	A
M33	C	G	G	G	C	T	T	G	G	T	G
	C	G	G	A	C	C	C	G	C	C	A
M34	n/a	n/a	G	G	C	T	T	n/a	G	n/a	n/a
	n/a	n/a	G	G	C	C	C	n/a	G	n/a	n/a

III. Equatorial Guinea

	TAS1R1		*TAS1R3*					*TAS2R16*	*TAS2R38*		
	C329T	**G1114A**	**G13A**	**G740A**	**C2269T**	**-C1572T**	**-C1266T**	**G516T**	**C145G**	**C785T**	**G886A**
EG1	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG2	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG3	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	T	T	T	C	T	A
EG4	C	G	G	G	C	C	T	T	G	C	G
	C	G	G	A	C	T	T	T	C	C	A
EG5	C	G	G	G	C	C	C	G	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
EG6	C	G	G	G	C	C	T	T	C	C	G
	C	A	G	G	C	T	T	T	C	C	G
EG7	C	G	G	G	C	C	C	G	G	C	G
	C	G	G	G	C	T	T	T	C	C	A
EG8	C	G	G	G	C	C	T	T	G	C	G
	C	G	G	G	C	C	T	T	C	C	A
EG9	C	G	G	A	C	T	T	G	G	C	G
	C	G	G	A	C	T	T	G	C	C	A
EG10	C	G	G	G	C	C	C	G	G	C	A
	C	G	G	G	C	C	C	G	G	T	A
EG11	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	A	C	T	C	T	C	C	G
EG12	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	T	T	C	T	A
EG13	C	G	G	G	C	T	T	G	G	C	G
	C	G	G	A	C	T	T	G	C	T	A
EG14	C	G	G	G	C	C	C	G	G	C	A
	C	G	G	A	C	T	T	T	G	C	A
EG15	C	G	G	G	C	T	T	G	C	C	G
	C	G	G	A	C	T	T	G	C	C	G
EG16	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	A	C	T	T	T	C	C	A
EG17	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	T	A
EG18	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	T	T	C	C	A
EG19	C	G	G	G	C	T	T	T	C	C	G
	C	G	G	G	C	T	T	T	C	C	G
EG20	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	C	T	A
EG21	C	G	G	G	C	C	T	T	G	C	G
	C	G	G	G	C	C	T	T	C	C	A
EG22	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
EG23	C	G	G	G	C	C	T	T	G	C	G
	C	G	G	A	C	T	T	T	C	C	A
EG24	C	G	G	G	C	C	C	T	C	C	G
	T	G	G	G	C	C	C	T	C	C	G
EG25	C	G	G	G	C	C	T	G	G	C	G
	C	A	G	G	C	T	T	T	C	T	A
EG26	C	G	G	G	C	T	T	T	G	C	G
	C	A	G	A	C	T	T	T	C	C	A
EG27	C	G	G	G	C	C	T	T	G	C	G
	C	G	G	G	C	C	T	T	C	C	A

EG28	C	G	G	G	C	T	T	T	G	C	G
	C	A	G	A	C	T	T	T	C	T	A
EG29	C	G	G	G	C	C	T	G	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG30	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	A	C	T	T	T	C	C	A
EG31	C	G	G	G	C	C	C	G	G	C	A
	C	G	G	G	C	T	T	T	G	T	A
EG32	C	G	G	G	C	C	T	G	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG33	C	G	G	G	C	C	C	G	G	C	A
	C	G	G	A	C	T	T	T	G	T	A
EG34	C	G	G	G	C	T	T	G	G	C	G
	C	G	G	A	C	T	T	T	C	C	A
EG35	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	A	C	T	T	T	C	C	A
EG36	C	G	G	G	C	C	C	G	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG37	C	G	G	G	C	T	T	G	G	C	A
	C	G	G	A	C	T	T	T	G	T	A
EG38	C	G	G	G	C	C	C	T	G	C	G
	T	G	G	G	C	C	C	T	C	T	A
EG39	C	G	G	G	C	C	C	G	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
EG40	C	G	G	G	C	T	T	T	G	T	A
	C	G	G	A	T	T	T	T	G	T	A
EG41	C	G	G	A	C	T	T	G	G	C	G
	C	G	G	A	C	T	T	T	C	T	A
EG42	C	G	G	G	C	C	T	T	C	C	G
	C	G	G	G	C	T	T	T	C	C	G
EG43	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	T	T	T	C	C	G
EG44	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	T	T	C	C	A
EG45	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	T	A
EG46	C	G	G	G	C	C	C	G	C	C	G
	C	A	G	G	C	C	T	T	C	C	G
EG47	C	G	G	G	C	C	T	G	G	C	G
	C	G	G	A	C	T	T	G	C	C	A
EG48	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	T	T	T	G	T	A
EG49	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	A	C	T	C	T	C	C	G
EG50	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	T	T	T	C	C	A
EG51	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG52	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	T	T	C	C	G
EG53	C	G	G	G	C	C	T	G	G	C	G
	C	G	G	A	C	T	T	T	C	C	G
EG54	C	G	G	G	C	T	T	T	G	T	A
	C	G	G	A	T	T	T	T	G	T	A
EG55	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG56	C	G	G	A	C	T	T	T	G	C	G
	C	G	G	A	C	T	T	T	C	T	A
EG57	C	G	G	G	C	C	C	G	G	C	G
	C	G	G	G	C	C	C	T	C	C	A
EG58	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
EG59	C	G	G	G	C	C	T	G	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
EG60	C	G	G	G	C	C	T	G	G	T	A
	C	G	G	G	C	T	T	G	G	T	A
EG61	C	G	G	G	C	C	C	G	G	C	G
	C	G	G	A	C	T	T	G	C	T	A
EG62	C	G	G	G	C	T	T	T	G	C	A
	C	G	G	A	C	T	T	T	G	T	A
EG63	C	G	G	G	C	C	T	T	C	C	G
	C	G	G	G	C	T	T	T	C	C	G
EG64	C	G	G	A	C	T	T	T	G	C	G
	C	A	G	A	C	T	T	T	C	C	A
EG65	C	G	G	G	C	C	C	G	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
EG66	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	T	T	C	T	A
EG67	C	G	G	G	C	C	C	G	G	T	A

	C	G	G	G	C	T	T	T	G	T	A
EG68	C	G	G	G	C	T	T	G	C	C	G
	C	G	G	A	C	T	T	T	C	C	G
EG69	C	G	G	G	C	T	T	T	G	C	A
	C	G	G	A	T	T	T	T	G	T	A
EG70	C	G	G	G	C	T	T	G	C	C	G
	C	G	G	G	C	T	T	T	C	C	G
EG71	C	G	G	G	C	C	C	G	G	C	G
	C	G	G	A	C	C	T	T	C	T	A
EG72	C	G	G	G	C	C	C	G	C	C	G
	C	G	G	G	C	C	C	G	C	C	G
EG73	C	G	G	G	C	C	C	T	G	C	A
	T	G	G	A	C	T	T	T	G	T	A
EG74	C	G	G	G	C	C	C	G	C	C	G
	C	G	G	G	C	T	T	T	C	C	G
EG75	C	G	G	G	C	C	C	G	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
EG76	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	C	T	A

IV. Uganda

	TAS1R1		*TAS1R3*					*TAS2R16*	*TAS2R38*		
	C329T	**G1114A**	**G13A**	**G740A**	**C2269T**	**-C1572T**	**-C1266T**	**G516T**	**C145G**	**C785T**	**G886A**
U1	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	C	T	C	C	A
U2	C	G	G	G	C	T	T	T	G	n/a	G
	C	G	G	A	C	C	T	T	C	n/a	A
U3	C	G	G	G	C	T	T	T	G	T	A
	C	G	G	G	C	T	T	T	G	C	A
U4	C	G	G	G	C	T	T	T	G	C	A
	C	G	G	G	C	C	C	G	G	C	A
U5	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
U6	C	G	G	G	C	T	T	T	G	T	A
	C	A	G	A	C	T	T	T	G	C	A
U7	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	C	T	C	C	A
U8	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	T	G	C	T	A
U9	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	T	T	C	C	A
U10	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	A	C	C	C	T	C	C	A
U11	C	G	G	G	C	T	T	T	G	T	A
	C	G	G	G	C	T	T	T	C	C	G
U12	C	G	G	G	C	C	C	T	G	T	G
	C	A	G	G	C	C	C	T	C	C	A
U13	n/a	G	G	G	C	T	T	n/a	G	n/a	n/a
	n/a	G	G	G	C	T	T	n/a	C	n/a	n/a
U14	C	G	G	G	C	C	C	T	G	C	A
	C	G	G	G	C	C	C	T	G	C	A
U15	C	G	G	G	C	C	T	T	G	T	A
	C	G	G	G	C	C	C	T	G	C	A
U16	C	G	G	G	T	T	T	T	G	T	G
	C	G	G	G	C	T	T	T	G	C	A
U17	C	G	G	A	C	T	T	T	G	C	G
	C	G	G	A	C	T	T	T	G	C	A
U18	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	T	T	T	C	C	A
U19	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	C	A
U20	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	T	G	C	C	A
U21	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	C	T	C	C	A
U22	C	G	G	G	C	T	T	T	C	C	G
	C	G	G	A	C	T	T	T	C	C	G
U23	C	G	A	G	C	C	T	T	C	C	G
	C	G	G	G	C	C	C	G	C	C	G
U24	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	T	T	T	C	C	A
U25	C	G	G	G	C	T	T	T	C	C	G
	C	G	G	G	C	C	T	T	C	C	G
U26	C	G	G	G	C	C	C	T	G	n/a	G
	C	G	G	G	C	C	C	G	C	n/a	A
U27	C	G	G	G	C	C	T	T	G	T	G

	C	G	G	G	C	C	T	T	C	C	A
U28	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	C	C	G	C	C	A
U29	C	G	G	G	C	T	T	T	C	C	G
	C	G	G	G	C	C	T	T	C	C	G
U30	C	G	G	G	C	T	T	T	G	T	A
	C	G	G	A	C	T	T	G	G	T	A
U31	C	G	G	A	C	T	T	T	C	C	G
	C	G	G	A	C	T	T	G	C	C	G
U32	C	G	G	G	C	T	T	T	G	T	T
	C	G	G	G	C	T	T	T	C	C	G
U33	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	C	A
U34	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	A	C	C	T	T	C	C	A
U35	C	G	G	G	C	C	T	T	G	T	G
	C	A	G	G	C	C	C	G	C	C	A
U36	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	T	T	T	C	C	A
U37	C	G	G	G	C	C	C	T	G	T	G
	C	G	G	G	C	C	C	T	C	C	A
U38	C	G	G	G	C	T	T	T	G	C	G
	C	A	G	G	C	T	T	T	C	C	A
U39	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	A	C	C	T	T	C	C	A
U40	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	T	T	T	C	C	A
U41	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	C	C	G	C	C	A
U42	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	C	T	T	C	C	A
U43	C	G	G	G	C	C	T	T	G	T	A
	C	G	G	G	C	C	C	G	G	T	A
U44	C	G	G	G	C	T	T	T	C	C	G
	C	G	G	G	C	T	T	G	C	C	G
U45	C	G	G	G	C	C	T	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
U46	C	G	G	G	C	T	T	T	G	C	G
	C	A	G	G	C	C	C	T	C	C	A
U47	C	G	G	G	C	T	T	T	G	T	A
	C	G	G	G	C	T	T	T	G	C	A
U48	C	G	G	G	C	T	T	T	G	T	G
	C	G	G	G	C	T	T	T	C	C	A
U49	C	n/a	G	G	C	T	T	G	n/a	C	G
	C	n/a	G	G	C	T	T	G	n/a	C	G

V. Portugal

	TAS1R1		***TAS1R3***					***TAS2R16***	***TAS2R38***		
	C329T	**G1114A**	**G13A**	**G740A**	**C2269T**	**-C1572T**	**-C1266T**	**G516T**	**C145G**	**C785T**	**G886A**
P1	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	C	G
P2	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P3	C	G	G	G	C	C	C	T	G	T	A
	C	A	G	G	C	C	C	T	G	T	A
P4	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
P5	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	C	T	A
P6	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	C	C	G
P7	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P8	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	C	T	A
P9	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P10	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P11	C	G	G	G	C	C	C	T	G	C	A
	C	A	G	G	T	T	T	T	G	C	A
P12	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	T	T	C	T	A
P13	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P14	C	G	G	G	C	T	T	T	G	C	G

	C	G	G	G	C	T	T	T	C	T	A
P15	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P16	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	T	T	C	T	A
P17	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P18	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	A	C	T	T	T	C	T	A
P19	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	C	T	A
P20	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P21	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P22	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	T	A
P23	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P24	C	G	G	G	C	C	C	T	G	C	G
	T	G	G	G	T	T	T	T	C	T	A
P25	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P26	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
P27	C	G	G	G	C	C	T	T	C	C	G
	C	A	G	G	C	T	T	T	C	C	G
P28	C	G	G	G	C	T	T	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
P29	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P30	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	T	A
P31	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	C	T	A
P32	C	G	G	G	C	C	C	T	C	C	G
	T	G	G	G	C	C	C	T	C	C	G
P33	C	G	G	G	C	C	C	T	C	T	A
	C	G	G	G	C	C	C	T	C	T	A
P34	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	T	A
P35	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	C	T	G	T	A
P36	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P37	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P38	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P39	C	G	G	G	C	C	C	T	G	C	G
	C	A	G	G	C	C	C	T	G	T	A
P40	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	T	T	T	T	C	T	A
P41	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	T	T	T	T	C	T	A
P42	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	C	T	A
P43	C	G	G	G	C	C	C	T	G	T	A
	C	G	G	G	C	C	T	T	G	T	A
P44	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	C	C	T	G	T	A
P45	C	G	G	G	C	C	C	T	G	C	G
	C	G	G	G	C	T	T	T	C	T	A
P46	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	T	T	T	T	C	C	G
P47	C	G	G	G	C	C	C	T	C	C	G
	C	G	G	G	C	C	C	T	C	C	G
P48	C	G	G	G	C	C	T	T	G	C	G
	C	G	G	G	T	T	T	T	C	T	A
P49	C	G	G	G	C	C	C	T	G	T	A
	C	A	G	G	C	C	C	T	G	T	A

Annex 2. *TAS2R16* supplementary material

Table S3 – Data and information collected for G516T from *TAS2R16*.

Paper/ Database	Publication/ Sample reference	Population	Data about the samples	Abbreviation	n	T516 frequency	Lifestyle
Soranzo *et al.* 2005	Curr Biol	Tunisian	Tunisia	Tun	17	1	n/a
Soranzo *et al.* 2005	Curr Biol	Bantu	South Africa	BSA	28	0.79	agriculture
Soranzo *et al.* 2005	Curr Biol	Amhara	Ethiopia	AE	20	0.925	agriculture
Soranzo *et al.* 2005	Curr Biol	Cameroonian	Cameroon	Ca	12	0.67	agriculture
HapMap	PhaseII+III, August 10	CEU	CEPH (Utah residents with ancestry from northern and western Europe)	-	65	1.000	n/a
HapMap	PhaseII+III, August 10	CHB	Han Chinese in Beijing, China	-	45	1.000	n/a
HapMap	PhaseII+III, August 10	JPT	Japanese in Tokyo, Japan	-	44	1.000	n/a
HapMap	PhaseII+III, August 10	YRI	Yoruba in Ibadan, Nigeria	-	147	0.633	agriculture
HapMap	PhaseII+III, August 10	ASW	African ancestry in Southwest USA	-	54	0.676	n/a
HapMap	PhaseII+III, August 10	LWK	Luhya in Webuye, Kenya	-	109	0.807	agriculture
HapMap	PhaseII+III, August 10	MEX	Mexican ancestry in Los Angeles, California	-	58	0.966	n/a
HapMap	PhaseII+III, August 10	MKK	Maasai in Kinyawa, Kenya	-	156	0.772	pastoralist
Hinrichs *et al.* 2006	Am J Hum Genet	Biaka Pygmy	Central African Republic	BP	36	0.79	hunter-gatherer
Hinrichs *et al.* 2006	Am J Hum Genet	Mandenka	Senegal	MS	24	0.79	agriculture
Hinrichs *et al.* 2006	Am J Hum Genet	Mbuti Pygmies	Democratic Republic of Congo	MP	15	0.8	hunter-gatherer
Hinrichs *et al.* 2006	Am J Hum Genet	Bantu NE	North-Eastern Africa	BNE	12	0.83	agriculture
Hinrichs *et al.* 2006	Am J Hum Genet	San	South	-	7	0.86	hunter-gatherer
Hinrichs *et al.* 2006	Am J Hum Genet	Mozabite	Algeria	MA	30	0.9	pastoralist
Hinrichs *et al.* 2006	Am J Hum Genet	Palestinian	Israel	Plt	50	0.94	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Maya	México	-	25	0.96	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Bedouin	North Africa	Be	47	0.98	pastoralist
Hinrichs *et al.* 2006	Am J Hum Genet	Brahui	Pakistan	Br	24	0.98	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Adygei	Russian Caucasus	Ad	17	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Balochi	Iran	Ba	26	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Makrani	Pakistan	Mak	26	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Sindhi	Pakistan	Si	24	1	agriculture
Hinrichs *et al.* 2006	Am J Hum Genet	Burusho	Pakistan	Bu	25	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Hazara	Pakistan	Ha	22	1	pastoralist
Hinrichs *et al.* 2006	Am J Hum Genet	Kalash	Pakistan	Ka	25	1	agriculture
Hinrichs *et al.* 2006	Am J Hum Genet	Pathan	Afghanistan	Pa	25	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Basque	Spain	Bsq	30	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	French	France	Fr	24	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Bergamo	Italy	Brg	27	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Sardinian	Italy	Sa	14	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Tuscan	Italy	Tus	8	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Cambodian	Cambodia	Cam	11	1	n/a

Hinrichs *et al.* 2006	Am J Hum Genet	Colombia	Colombia	Co	13	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Dai	China	-	10	1	agriculture
Hinrichs *et al.* 2006	Am J Hum Genet	Lahu	China	Lh	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Naxi	China	Nx	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Daur	China	Dr	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Druze	Levant/East mediterranean	Drz	47	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Hezhen	Orient/China	He	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Oroqen	northern China	Orq	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Tu	China/Tibete	-	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Karitiana	Brazil	Kar	24	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Miaozu	China	Mi	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Uygur	Central Asia	Uy	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Xibo	China	Xb	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Mongola	China	Mo	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Melanesian	Melanesia (Oceania)	Me	22	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Orcadian	North Europe	Orc	16	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Papuan	New Guinea	Pp	17	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Pima	Arizona USA	-	25	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Russian	Russia	Ru	25	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	She	China	-	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Tujia	China	Tj	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Yizu	China	Yi	10	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Surui	Brazil	-	21	1	n/a
Hinrichs *et al.* 2006	Am J Hum Genet	Yakut	Turkey	Ykt	25	1	n/a
ALFRED	SA000100B	Hausa	Zaria, North Central Nigeria	Hs	39	0.73	agriculture
ALFRED	SA000099S	Ibo	Enugu, southeastern Nigeria	-	48	0.576	agriculture
ALFRED	SA002770Q	Lisongo	Central Africa	-	8	0.62	hunter-gatherer(?)
ALFRED	SA000487T	Chagga	Dar-es-Salaam, Tanzania; born in Mount Kilimanjaro region	Ch	45	0.674	agriculture
ALFRED	SA000015G	Ethiopian Jews	Israel (?)	EJ	32	0.95	n/a
ALFRED	SA001773S	Sandawe	North Central Tanzania, Dodoma region	Sdw	40	0.925	hunter-gatherer
ALFRED	SA002586V	Zaramo	Tanzania, mostly around Dar es Salaam, Bagamoyo	Zrm	39	0.81	agriculture
ALFRED	SA000101C	African American	USA (from NIGMS Cell Repository, Corielle Institute for Medical Research)	AA	89	0.753	n/a
ALFRED	SA000016H	Yemenite Jews	born in Yemen	YJ	42	0.988	n/a
ALFRED	SA002765U	Kuwaiti	Arabs from Kuwait	Kw	16	1	n/a
ALFRED	SA004036N	Tsaatan	Tsaatan (Reindeer people) from near lake Baikal region	Ts	42	0.976	n/a

Table S4 – AMOVA groups regarding *TAS2R16* containing only African populations; **I.** Lifestyle, **II.** Malaria risk, **III.** African Region.

I. Lifestyle

Agriculture		Hunting-Gathering	Pastoralism
Mozambique	YRI	Biaka Pygmies	Tunisian
Angola	Zaramo	Mbuti Pygmies	Mozabite
Bantu SA	Chagga	Sandawe	Uganda
LWK	Ibo	San	MKK
Mandenka	Hausa		Bedouin
Cameroonian	Equatorial Guinea		
Amhara	Bantu NE		

II. Malaria Risk

Everywhere		Intermediate	Not known
Mozambique	Equatorial Guinea	MKK	Bantu SA
Biaka Pygmies	Uganda	Amhara	Tunisian
Mbuti Pygmies	Hausa	LWK	Mozabite
Lisongo	YRI	Zaramo	Bedouin
Angola	Ibo	Chagga	
Cameroonian	Mandenka	Sandawe	
		San	

III. African Region

South	Central	North	West	East	South-East
Angola	Biaka Pygmies	Tunisian	Mandenka	Amhara	Mozambique
San	Mbuti Pygmies	Mozabite	YRI	LWK	
Bantu SA	Lisongo		Cameroonian	MKK	
			Equatorial Guinea	Uganda	
			Hausa	Chagga	
			Ibo	Sandawe	
				Zaramo	
				Bantu NE	

Table S5 – AMOVA groups regarding *TAS2R16* containing all populations; **I.** Lifestyle, **II.** Malaria risk, **III.** Region, **IV.** Continent.

I. Lifestyle

Agriculture		Hunting-Gathering	Pastoralism
Mozambique	Bantu NE	Biaka Pygmies	Tunisian
Angola	Sindhi	Mbuti Pygmies	Mozabite
Bantu SA	Pathan	Sandawe	Uganda
LWK	Kalash	San	MKK
Mandenka	Dai	Hazara	Bedouin
Cameroonian	Equatorial Guinea	Balochi	
Amhara	Hausa		
YRI	Ibo		
Zaramo	Chagga		

II. Malaria Risk

Everywhere	Intermediate		Not known	
Mozambique	MKK	Miaozu	Bantu SA	African American
Biaka Pygmies	Amhara	Yizu	Tunisian	Tsaatan
Mbuti Pygmies	LWK	She	Mozabite	Sardinian
Lisongo	Zaramo	Tujia	Bedouin	Tuscan
Angola	Chagga	Papuan	CEU	Bergamo
Cameroonian	Sandawe	Makrani	JPT	Pima
Equatorial Guinea	San	Pathan	Portugal	Adygei
Uganda	Colombia	Sindhi	Basque	Yakut
Hausa	Maya	Burusho	French	Palestinian
YRI	CHB	Hazara	Orcadian	Kuwaiti
Ibo	Dai	Kalash	Russian	Druze
Mandenka	Lahu	Balochi	ASW	
Karitiana	Naxi	Brahui		
Surui	Daur	Yemenite Jews		
Cambodian	Oroqen	Hezhen		
	Tu	Uygur		
	Xibo	Ethiopian Jews		
	Mongola			

III. Region

Group1	Group2	Group3	Group4	Group5	Group6
Angola	Biaka Pygmies	Tunisian	Mandenka	Bantu NE	Mozambique
San	Mbuti Pygmies	Mozabite	YRI	Amhara	
Bantu SA	Lisongo		Cameroonian	LWK	
			Equatorial Guinea	MKK	
			Hausa	Uganda	
			Ibo	Chagga	
				Sandawe	
				Zaramo	

Group7	Group8	Group9	Group10	Group11	Group12
MEX	Colombia	JPT	She	CHB	Hezhen
Pima	Karitiana			Tujia	
Maya	Surui				

Group13	Group14	Group15	Group16	Group17	Group18
Naxi	Daur	Yakut	Dai	Adygei	Bergamo
	Mongola		Lahu		Sardinian
	Oroqen		Miaozu		Tuscan
			Yizu		
			Cambodian		

Group19	Group20	Group21	Group22	Group23	Group24
Portugal	Russian	Bedouin	Druze	Melanesian	Hazara
Orcadian			Palestinian		Tu
Basque					Uygur
French					Xibo

Group25	Group26	Group27	Group28	Group29	Group30
Balochi	Papuan	Ethiopian Jews	Kuwaiti	Tsaatan	African American
Brahui					
Burusho					
Kalash					
Makrani					
Pathan					
Sindhi					

Group31	Group32	Group33
CEU	ASW	Yemenite Jews

IV. Continent

Africa		America		Asia		Europe	Oceania
Mandenka	Chagga	MEX	JPT	Pathan	Tujia	Bergamo	Melanesian
YRI	Sandawe	Pima	Hazara	Sindhi	Naxi	Sardinian	Papuan
Cameroonian	Zaramo	Maya	Tu	Dai	Daur	Tuscan	
Equatorial Guinea	Mozambique	Colombia	Uygur	Lahu	Mongola	Portugal	
Hausa	Biaka Pygmies	Karitiana	Xibo	Miaozu	Oroqen	Orcadian	
Ibo	Mbuti Pygmies	Surui	Balochi	Yizu	Druze	Basque	
Amhara	Lisongo		Brahui	Cambodian	Palestinian	French	
LWK	Tunisian		Burusho	Hezhen	Kuwaiti	Russian	
MKK	Mozabite		Kalash	She	Tsaatan	CEU	
Uganda	Angola		Makrani	CHB			
Bantu SA	San						

Table S6 – F_{ST} values obtained to *TAS2R16* and used to construct the MDS plot presented in chapter IV (Results & Discussion). This table was split in three parts and includes 75 populations among the world. MZ: Mozambique; AN: Angola; EG: Equatorial Guinea; UG: Uganda; PT: Portugal. The remaining abbreviations are equal to the described in table S3.

	MZ	BP	MP	AN	Ca	EG	UG	Hs	YRI	Ibo	MS	BSA	Tun	MA	Be	MKK	AE	LWK	Zrm	Ch	Sdw	San	BNE	Co
MZ	*																							
BP	0.0052	*																						
MP	-0.0016	-0.0240	*																					
AN	0.0072	0.0676	0.0598	*																				
Ca	-0.0253	0.0147	0.0081	-0.0216	*																			
EG	0.0054	-0.0103	-0.0197	0.0673	0.0133	*																		
UG	0.0561	0.0030	-0.0096	0.1509	0.0906	0.0093	*																	
Hs	-0.0124	-0.0033	-0.0106	0.0217	-0.0176	-0.0023	0.0378	*																
YRI	0.0027	0.0465	0.0415	-0.0086	-0.0205	0.0458	0.1036	0.0130	*															
Ibo	0.0249	0.0910	0.0835	-0.0121	-0.0081	0.0917	0.1723	0.0417	0.0006	*														
MS	0.0013	-0.0177	-0.0276	0.0615	0.0099	-0.0139	*	-0.0070	0.0424	0.0847	*													
BSA	0.0002	-0.0160	-0.0256	0.0586	0.0076	-0.0126	0.0041	-0.0074	0.0401	0.0818	-0.0196	*												
Tun	0.2214	0.1441	0.1850	0.3273	0.3539	0.1313	0.0832	0.1927	0.2340	0.3230	0.1626	0.1601	*											
MA	0.0958	0.0281	0.0178	0.1985	0.1517	0.0332	-0.0052	0.0733	0.1384	0.2136	0.0272	0.0319	0.0585	*										
Be	0.2635	0.1633	0.2103	0.3950	0.4281	0.1397	0.0827	0.2250	0.2435	0.3741	0.1885	0.1856	-0.0055	0.0480	*									
MKK	0.0032	-0.0075	-0.0164	0.0621	0.0085	-0.0048	0.0150	-0.0032	0.0427	0.0879	-0.0111	-0.0101	0.1275	0.0385	0.1262	*								
AE	0.1157	0.0452	0.0394	0.2168	0.1815	0.0488	0.0036	0.0924	0.1564	0.2299	0.0463	0.0506	0.0438	-0.0174	0.0219	0.0536	*							
LWK	0.0204	-0.0085	-0.0192	0.0977	0.0360	-0.0040	0.0011	0.0087	0.0673	0.1242	-0.0121	-0.0099	0.1070	0.0191	0.1066	-0.0003	0.0324	*						
Zrm	0.0145	-0.0127	-0.0234	0.0847	0.0289	-0.0082	-0.0027	0.0038	0.0587	0.1084	-0.0163	-0.0141	0.1280	0.0180	0.1416	-0.0044	0.0335	-0.0088	*					
Ch	-0.0112	0.0200	0.0136	-0.0022	-0.0268	0.0198	0.0774	-0.0053	-0.0028	0.0125	0.0158	0.0142	0.2319	0.1166	0.2703	0.0166	0.1352	0.0386	0.0311	*				
Sdw	0.1413	0.0596	0.0554	0.2576	0.2222	0.0608	0.0116	0.1136	0.1717	0.2647	0.0624	0.0668	0.0346	-0.0108	0.0210	0.0631	-0.0191	0.0416	0.0459	0.1610	*			
San	0.0145	-0.0303	-0.0437	0.0910	0.0349	-0.0235	-0.0426	0.0000	0.0678	0.1167	-0.0339	-0.0300	0.1826	-0.0360	0.1470	-0.0171	-0.0220	-0.0309	-0.0354	0.0343	-0.0130	*		
BNE	0.0135	-0.0229	-0.0351	0.0854	0.0311	-0.0171	-0.0243	0.0009	0.0621	0.1096	-0.0265	-0.0234	0.1651	-0.0084	0.1705	-0.0119	0.0092	-0.0214	-0.0258	0.0315	0.0207	-0.0577	*	
Co	0.2034	0.1299	0.1590	0.3040	0.3150	0.1220	0.0736	0.1771	0.2267	0.3063	0.1436	0.1428	0.0000	0.0478	-0.0110	0.1211	0.0321	0.0999	0.1151	0.2165	0.0265	0.1462	0.1379	*
Maya	0.1778	0.0961	0.1087	0.2897	0.2838	0.0913	0.0388	0.1492	0.1995	0.2905	0.1054	0.1074	0.0094	0.0082	-0.0090	0.0909	-0.0111	0.0691	0.0806	0.1930	-0.0058	0.0426	0.0733	0.0016
CHB	0.3174	0.2187	0.3195	0.4484	0.5328	0.1766	0.1285	0.2754	0.2701	0.4130	0.2630	0.2516	0.0000	0.1097	0.0101	0.1538	0.1005	0.1377	0.1945	0.3146	0.0683	0.3614	0.3057	0.0000
Dai	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Lh	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Nx	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Dr	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Orq	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Tu	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Xb	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Mo	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Mi	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Yi	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
She	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Tj	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Pp	0.2214	0.1441	0.1850	0.3273	0.3539	0.1313	0.0832	0.1927	0.2340	0.3230	0.1626	0.1601	0.0000	0.0585	-0.0055	0.1275	0.0438	0.1070	0.1280	0.2319	0.0346	0.1826	0.1651	0.0000
Mak	0.2563	0.1711	0.2349	0.3725	0.4254	0.1480	0.1000	0.2226	0.2472	0.3557	0.1991	0.1933	0.0000	0.0774	0.0017	0.1377	0.0646	0.1188	0.1521	0.2618	0.0476	0.2513	0.2175	0.0000
Pa	0.2526	0.1683	0.2298	0.3678	0.4183	0.1463	0.0983	0.2195	0.2459	0.3523	0.1953	0.1898	0.0000	0.0755	0.0011	0.1367	0.0625	0.1176	0.1496	0.2587	0.0464	0.2443	0.2121	0.0000
Si	0.2490	0.1655	0.2246	0.3631	0.4110	0.1446	0.0966	0.2163	0.2445	0.3488	0.1915	0.1864	0.0000	0.0736	0.0005	0.1357	0.0604	0.1164	0.1471	0.2555	0.0451	0.2372	0.2066	0.0000
Bu	0.2526	0.1683	0.2298	0.3678	0.4183	0.1463	0.0983	0.2195	0.2459	0.3523	0.1953	0.1898	0.0000	0.0755	0.0011	0.1367	0.0625	0.1176	0.1496	0.2587	0.0464	0.2443	0.2121	0.0000
Ha	0.2414	0.1597	0.2138	0.3534	0.3959	0.1410	0.0930	0.2099	0.2417	0.3418	0.1836	0.1792	0.0000	0.0696	-0.0010	0.1336	0.0559	0.1140	0.1419	0.2491	0.0423	0.2226	0.1954	0.0000
Ka	0.2526	0.1683	0.2298	0.3678	0.4183	0.1463	0.0983	0.2195	0.2459	0.3523	0.1953	0.1898	0.0000	0.0755	0.0011	0.1367	0.0625	0.1176	0.1496	0.2587	0.0464	0.2443	0.2121	0.0000
Ba	0.2563	0.1711	0.2349	0.3725	0.4254	0.1480	0.1000	0.2226	0.2472	0.3557	0.1991	0.1933	0.0000	0.0774	0.0017	0.1377	0.0646	0.1188	0.1521	0.2618	0.0476	0.2513	0.2175	0.0000
Br	0.2090	0.1259	0.1554	0.3220	0.3378	0.1148	0.0629	0.1787	0.2202	0.3166	0.1416	0.1409	-0.0075	0.0323	-0.0160	0.1111	0.0106	0.0901	0.1091	0.2208	0.0115	0.1069	0.1239	-0.0139
YJ	0.2765	0.1790	0.2422	0.4055	0.4564	0.1507	0.0973	0.2378	0.2509	0.3810	0.2097	0.2041	-0.0125	0.0671	-0.0087	0.1354	0.0441	0.1168	0.1571	0.2804	0.0358	0.2082	0.2107	-0.0178
He	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
Uy	0.1877	0.1170	0.1367	0.2841	0.2814	0.1129	0.0641	0.1634	0.2196	0.2920	0.1270	0.1276	0.0000	0.0374	-0.0177	0.1144	0.0208	0.0925	0.1032	0.2030	0.0181	0.1149	0.1144	0.0000
CEU	0.3708	0.2618	0.3894	0.5101	0.6092	0.2031	0.1549	0.3226	0.2914	0.4626	0.3189	0.3033	0.0000	0.1393	0.0162	0.1683	0.1334	0.1549	0.2332	0.3616	0.0870	0.4462	0.3784	0.0000
JPT	0.3145	0.2164	0.3155	0.4449	0.5282	0.1752	0.1272	0.2728	0.2689	0.4103	0.2599	0.2488	0.0000	0.1081	0.0097	0.1530	0.0987	0.1368	0.1924	0.3121	0.0673	0.3564	0.3016	0.0000
PT	0.3289	0.2278	0.3348	0.4620	0.5504	0.1821	0.1340	0.2854	0.2745	0.4237	0.2750	0.2626	0.0000	0.1158	0.0114	0.1568	0.1073	0.1412	0.2026	0.3246	0.0722	0.3804	0.3216	0.0000
Bsq	0.2703	0.1819	0.2547	0.3903	0.4521	0.1545	0.1065	0.2346	0.2524	0.3688	0.2137	0.2066	0.0000	0.0848	0.0039	0.1415	0.0728	0.1231	0.1617	0.2738	0.0525	0.2777	0.2381	0.0000
Fr	0.2490	0.1655	0.2246	0.3631	0.4110	0.1446	0.0966	0.2163	0.2445	0.3488	0.1915	0.1864	0.0000	0.0736	0.0005	0.1357	0.0604	0.1164	0.1471	0.2555	0.0451	0.2372	0.2066	0.0000
Orc	0.2171	0.1408	0.1788	0.3217	0.3447	0.1292	0.0810	0.1890	0.2323	0.3190	0.1581	0.1560	0.0000	0.0561	-0.0067	0.1261	0.0411	0.1054	0.1250	0.2283	0.0328	0.1739	0.1586	0.0000
Ru	0.2526	0.1683	0.2298	0.3678	0.4183	0.1463	0.0983	0.2195	0.2459	0.3523	0.1953	0.1898	0.0000	0.0755	0.0011	0.1367	0.0625	0.1176	0.1496	0.2587	0.0464	0.2443	0.2121	0.0000
ASW	-0.0100	0.0215	0.0153	-0.0018	-0.0259	0.0213	0.0777	-0.0040	-0.0023	0.0127	0.0173	0.0157	0.2257	0.1154	0.2582	0.0182	0.1336	0.0400	0.0325	-0.0103	0.1573	0.0360	0.0331	0.2121
AA	-0.0045	-0.0056	-0.0136	0.0422	-0.0044	-0.0038	0.0245	-0.0080	0.0281	0.0652	-0.0092	-0.0089	0.1498	0.0523	0.1584	-0.0034	0.0688	0.0036	-0.0008	0.0058	0.0818	-0.0089	-0.0059	0.1411
Ts	0.2468	0.1502	0.1902	0.3749	0.4013	0.1306	0.0744	0.2102	0.2355	0.3581	0.1721	0.1702	-0.0038	0.0401	-0.0113	0.1201	0.0151	0.0999	0.1298	0.2554	0.0158	0.1258	0.1510	-0.0096
Sa	0.2058	0.1318	0.1625	0.3070	0.3202	0.1233	0.0749	0.1792	0.2277	0.3085	0.1462	0.1452	0.0000	0.0493	-0.0102	0.1220	0.0337	0.1009	0.1168	0.2185	0.0277	0.1510	0.1415	0.0000
Tus	0.1753	0.1063	0.1193	0.2690	0.2558	0.1049	0.0556	0.1523	0.2133	0.2809	0.1137	0.1153	0.0000	0.0281	-0.0247	0.1080	0.0109	0.0855	0.0932	0.1922	0.0100	0.0907	0.0961	0.0000
Brg	0.2598	0.1738	0.2400	0.3771	0.4323	0.1496	0.1017	0.2256	0.2485	0.3590	0.2028	0.1967	0.0000	0.0793	0.0023	0.1387	0.0667	0.1199	0.1545	0.2648	0.0489	0.2581	0.2227	0.0000
Pima	0.2526	0.1683	0.2298	0.3678	0.4183	0.1463	0.0983	0.2195	0.2459	0.3523	0.1953	0.1898	0.0000	0.0755	0.0011	0.1367	0.0625	0.1176	0.1496	0.2587	0.0464	0.2443	0.2121	0.0000
Ad	0.2214	0.1441	0.1850	0.3273	0.3539	0.1313	0.0832	0.1927	0.2340	0.3230	0.1626	0.1601	0.0000	0.0585	-0.0055	0.1275	0.0438	0.1070	0.1280	0.2319	0.0346	0.1826	0.1651	0.0000
Ykt	0.2526	0.1683	0.2298	0.3678	0.4183	0.1463	0.0983	0.2195	0.2459	0.3523	0.1953	0.1898	0.0000	0.0755	0.0011	0.1367	0.0625	0.1176	0.1496	0.2587	0.0464	0.2443	0.2121	0.0000
Plt	0.1789	0.0866	0.0897	0.3051	0.2803	0.0830	0.0273	0.1469	0.1959	0.3048	0.0934	0.0972	0.0219	-0.0019	0.0086	0.0814	-0.0159	0.0593	0.0701	0.1968	-0.0096	0.0114	0.0490	0.0151
Kw	0.2171	0.1408	0.1788	0.3217	0.3447	0.1292	0.0810	0.1890	0.2323	0.3190	0.1581	0.1560	0.0000	0.0561	-0.0067	0.1261	0.0411	0.1054	0.1250	0.2283	0.0328	0.1739	0.1586	0.0000
Drz	0.3232	0.2233	0.3272	0.4553	0.5417	0.1794	0.1313	0.2805	0.2723	0.4184	0.2690	0.2571	0.0000	0.1128	0.0108	0.1553	0.1039	0.1395	0.1986	0.3196	0.0703	0.3710	0.3137	0.0000
Kar	0.2490	0.1655	0.2246	0.3631	0.4110	0.1446	0.0966	0.2163	0.2445	0.3488	0.1915	0.1864	0.0000	0.0736	0.0005	0.1357	0.0604	0.1164	0.1471	0.2555	0.0451	0.2372	0.2066	0.0000
Surui	0.2376	0.1567	0.2083	0.3484	0.3880	0.1392	0.0912	0.2066	0.2403	0.3382	0.1796	0.1756	0.0000	0.0675	-0.0017	0.1325	0.0536	0.1127	0.1393	0.2458	0.0409	0.2150	0.1896	0.0000
Cam	0.1932	0.1216	0.1445	0.2910	0.2932	0.1162	0.0676	0.1683	0.2222	0.2970	0.1329	0.1330	0.0000	0.0412	-0.0151	0.1169	0.0250	0.0952	0.1075	0.2078	0.0212	0.1258	0.1226	0.0000
MEX	0.2479	0.1434	0.1713	0.3837	0.3921	0.1258	0.0646	0.2093	0.2350	0.3680	0.1619	0.1621	0.0038	0.0269	-0.0066	0.1145	0.0022	0.0937	0.1220	0.2588	0.0063	0.0866	0.1247	-0.0020
Me	0.2414	0.1597	0.2138	0.3534	0.3959	0.1410	0.0930	0.2099	0.2417	0.3418	0.1836	0.1792	0.0000	0.0696	-0.0010	0.1336	0.0559	0.1140	0.1419	0.2491	0.0423	0.2226	0.1954	0.0000
EJ	0.1796	0.0940	0.1042	0.2967	0.2868	0.0890	0.0354	0.1496	0.1983	0.2960	0.1030	0.1054	0.0142	0.0047	-0.0025	0.0879	-0.0133	0.0661	0.0780	0.1952	-0.0074	0.0320	0.0663	0.0067

	Maya	CHB	Dai	Lh	Nx	Dr	Orq	Tu	Xb	Mo	Mi	Yi	She	Tj	Pp	Mak	Pa	Si	Bu	Ha	Ka	Ba	Br	YJ	He	Uy
Maya	*																									
CHB	0.0403	*																								
Dai	-0.0067	0.0000	*																							
Lh	-0.0067	0.0000	0.0000	*																						
Nx	-0.0067	0.0000	0.0000	0.0000	*																					
Dr	-0.0067	0.0000	0.0000	0.0000	0.0000	*																				
Orq	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	*																			
Tu	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*																		
Xb	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*																	
Mo	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*																
Mi	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*															
Yi	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*														
She	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*													
Tj	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*												
Pp	0.0094	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*											
Mak	0.0216	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*										
Pa	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*									
Si	0.0192	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*								
Bu	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*							
Ha	0.0167	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*						
Ka	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*					
Ba	0.0216	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*				
Br	-0.0145	0.0140	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0212	-0.0075	0.0017	0.0009	0.0000	0.0009	-0.0018	0.0009	0.0017	*			
YJ	0.0020	0.0008	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0242	-0.0125	-0.0060	-0.0065	-0.0071	-0.0065	-0.0084	-0.0065	-0.0060	-0.0139	*		
He	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0212	-0.0242	*	
Uy	-0.0067	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0212	-0.0242	0.0000	*
CEU	0.0569	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0242	0.0054	0.0000	0.0000
JPT	0.0394	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0134	0.0006	0.0000	0.0000
PT	0.0437	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0161	0.0019	0.0000	0.0000
Bsq	0.0260	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0047	-0.0041	0.0000	0.0000
Fr	0.0192	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0071	0.0000	0.0000
Orc	0.0077	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0088	-0.0136	0.0000	0.0000
Ru	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	-0.0065	0.0000	0.0000
ASW	0.1890	0.2982	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2000	0.2257	0.2518	0.2491	0.2463	0.2491	0.2407	0.2491	0.2518	0.2141	0.2674	0.2000	0.2000
AA	0.1116	0.1916	0.1327	0.1327	0.1327	0.1327	0.1327	0.1327	0.1327	0.1327	0.1327	0.1327	0.1327	0.1327	0.1498	0.1652	0.1636	0.1621	0.1636	0.1588	0.1636	0.1652	0.1342	0.1682	0.1327	0.1327
Ts	-0.0116	0.0133	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0164	-0.0038	0.0039	0.0032	0.0025	0.0032	0.0010	0.0032	0.0039	-0.0164	-0.0080	-0.0164	-0.0164
Sa	0.0027	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0129	-0.0169	0.0000	0.0000
Tus	-0.0148	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0287	-0.0311	0.0000	0.0000
Brg	0.0227	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0025	-0.0055	0.0000	0.0000
Pima	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	-0.0065	0.0000	0.0000
Ad	0.0094	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0075	-0.0125	0.0000	0.0000
Ykt	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	-0.0065	0.0000	0.0000
Plt	-0.0112	0.0471	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0219	0.0322	0.0312	0.0302	0.0312	0.0281	0.0312	0.0322	0.0015	0.0204	0.0076	0.0076
Kw	0.0077	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0088	-0.0136	0.0000	0.0000
Drz	0.0420	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0151	0.0014	0.0000	0.0000
Kar	0.0192	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0071	0.0000	0.0000
Surui	0.0154	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0028	-0.0091	0.0000	0.0000
Cam	-0.0036	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0184	-0.0217	0.0000	0.0000
MEX	-0.0141	0.0211	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	-0.0088	0.0038	0.0115	0.0108	0.0101	0.0108	0.0086	0.0108	0.0115	-0.0118	0.0001	-0.0088	-0.0088
Me	0.0167	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0018	-0.0084	0.0000	0.0000
EJ	-0.0176	0.0431	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	-0.0013	0.0142	0.0257	0.0246	0.0235	0.0246	0.0211	0.0246	0.0257	-0.0086	0.0094	-0.0013	-0.0013

	CEU	JPT	PT	Bsq	Fr	Orc	Ru	ASW	AA	Ts	Sa	Tus	Brg	Pima	Ad	Ykt	Plt	Kw	Drz	Kar	Surui	Cam	MEX	Me	EJ
CEU	*																								
JPT	0.0000	*																							
PT	0.0000	0.0000	*																						
Bsq	0.0000	0.0000	0.0000	*																					
Fr	0.0000	0.0000	0.0000	0.0000	*																				
Orc	0.0000	0.0000	0.0000	0.0000	0.0000	*																			
Ru	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	*																		
ASW	0.3401	0.2959	0.3070	0.2623	0.2463	0.2224	0.2491	*																	
AA	0.2161	0.1903	0.1967	0.1712	0.1621	0.1478	0.1636	0.0073	*																
Ts	0.0206	0.0129	0.0149	0.0063	0.0025	-0.0051	0.0032	0.2446	0.1498	*															
Sa	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2139	0.1423	-0.0087	*														
Tus	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1902	0.1251	-0.0235	0.0000	*													
Brg	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2544	0.1667	0.0045	0.0000	0.0000	*												
Pima	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2491	0.1636	0.0032	0.0000	0.0000	0.0000	*											
Ad	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2257	0.1498	-0.0038	0.0000	0.0000	0.0000	0.0000	*										
Ykt	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2491	0.1636	0.0032	0.0000	0.0000	0.0000	0.0000	0.0000	*									
Plt	0.0601	0.0464	0.0498	0.0358	0.0302	0.0204	0.0312	0.1906	0.1047	0.0047	0.0161	0.0001	0.0331	0.0312	0.0219	0.0312	*								
Kw	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2224	0.1478	-0.0051	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0204	*							
Drz	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3026	0.1941	0.0141	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0485	0.0000	*						
Kar	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2463	0.1621	0.0025	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0302	0.0000	0.0000	*					
Surui	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2378	0.1571	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0270	0.0000	0.0000	0.0000	*				
Cam	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2043	0.1358	-0.0138	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0105	0.0000	0.0000	0.0000	0.0000	*			
MEX	0.0286	0.0207	0.0227	0.0139	0.0101	0.0025	0.0108	0.2477	0.1464	-0.0084	-0.0011	-0.0158	0.0121	0.0108	0.0038	0.0108	-0.0020	0.0025	0.0219	0.0101	0.0077	-0.0061	*		
Me	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2407	0.1588	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0281	0.0000	0.0000	0.0000	0.0000	0.0000	0.0086	*	
EJ	0.0585	0.0423	0.0463	0.0298	0.0235	0.0125	0.0246	0.1895	0.1097	-0.0057	0.0078	-0.0092	0.0268	0.0246	0.0142	0.0246	-0.0113	0.0125	0.0447	0.0235	0.0198	0.0017	-0.0102	0.0211	*

The significant values of genetic distances after Bonferroni's correction are labelled in red.

Annex 3. *TAS2R38* supplementary material

Table S7 – Data and information collected for *TAS2R38* gene.

Paper	Publication	Population	Data about Sampling	Abbreviation	n	PAV	AAV	AAI	AVI	PVI	PAI	Lifestyle
Wooding *et al.* 2004	Am J Hum Genet	African	9 sub-Saharan Africans from Coriell Human Variation panel HD12, 22 Cameroonians	SSA	31	0.6129	0.0323	0.1774	0.1613	0.0161	0	n/a
Wooding *et al.* 2004	Am J Hum Genet	European	(10 hungarians, 45 utah samples from Centre d'Etude du Polymorphisme Humain)	NWE	55	0.4636	0.0455	0	0.4909	0	0	n/a
Wooding *et al.* 2004	Am J Hum Genet	North-American	(10 southwest Native Americans)	NA	10	0.95	0	0	0.05	0	0	n/a
Wooding *et al.* 2010	Chem Senses	Caucasian	Texas	Cc	50	0.49	0.06	0	0.49	0.04	0.02	n/a
Campbell *et al.* 2012	Mol Biol Evol	Afroasiatic	Cameroon	AA_WCA	73	0.506	0.007	0.22	0.267	0	0	Agriculture/ pastoralism
Campbell *et al.* 2012	Mol Biol Evol	Nilo-Saharan	Cameroon	NS_WCA	26	0.4031	0.0193	0.2123	0.3653	0	0	Agriculture
Campbell *et al.* 2012	Mol Biol Evol	Niger-Kordofanian	Cameroon	NK_WCA	62	0.387	0.008	0.307	0.298	0	0	Agriculture
Campbell *et al.* 2012	Mol Biol Evol	Fulani	Cameroon	-	48	0.553	0	0.145	0.308	0	0	Pastoralism
Campbell *et al.* 2012	Mol Biol Evol	Pygmy	Cameroon	-	62	0.54	0	0.25	0.21	0	0	Hunting and Gathering (HG)
Campbell *et al.* 2012	Mol Biol Evol	Afroasiatic	Kenya	AA_EA	132	0.492	0.038	0.152	0.318	0	0	Agriculture/ pastoralism/HG
Campbell *et al.* 2012	Mol Biol Evol	Nilo-Saharan	Kenya	NS_EA	130	0.568	0.023	0.1	0.309	0	0	Pastoralism (n=114) /HG (n=16)
Campbell *et al.* 2012	Mol Biol Evol	Niger-Kordofanian	Kenya	NK_EA	34	0.514	0.015	0.25	0.221	0	0	Agriculture
Campbell *et al.* 2012	Mol Biol Evol	Luo	Kenya	-	21	0.667	0	0.143	0.19	0	0	Pastoralism
Campbell *et al.* 2012	Mol Biol Evol	Hadza/Sandawe	Tanzania (Khoisan)	-	23	0.587	0	0.152	0.261	0	0	Hunting and Gathering
Campbell *et al.* 2012	Mol Biol Evol	Americas	Mexico/Nahua	-	13	0.846	0	0	0.154	0	0	n/a
Campbell *et al.* 2012	Mol Biol Evol	Middle East	Palestinian	Plt	20	0.5	0	0.05	0.45	0	0	n/a
Campbell *et al.* 2012	Mol Biol Evol	Pakistan	Brahui	Br	8	0.4375	0	0	0.5625	0	0	n/a
Campbell *et al.* 2012	Mol Biol Evol	East Asia	Chinese/Japanese	EAsian	71	0.627	0	0	0.373	0	0	n/a

Table S8 – AMOVA groups with significant percentages of variation among groups for *TAS2R38*.

I. Region

Group1	Group2	Group3	Group4	Group5	Group6
Afroasiatic WCA	Uganda	Angola	Mozambique	Nahua	Native American
Equatorial Guinea	Khoisan				
Nilo-Saharan WCA	Luo				
Niger-Kordofanian WCA	Niger-Kordofanian EA				
Fulani	Afroasiatic EA				
Pygmy	NiloSaharan EA				

Group7	Group8	Group9	Group10	Group11
Portugal	NW European	Brahui	Palestinian	East Asian

II. Continent

Africa		Europe	Asia	America
Angola	Pygmy	Portugal	Brahui	Nahua
Mozambique	Khoisan	NW European	Palestinian	Native American
Equatorial Guinea	Luo		East Asian	
Uganda	Niger-Kordofanian EA			
Afroasiatic WCA	Afroasiatic EA			
Nilo-Saharan WCA	Nilo-Saharan EA			
Niger-Kordofanian WCA	African			
Fulani				

Table S9 – F_{ST} values obtained to *TAS2R38* and used to construct the MDS plot presented in chapter IV (Results & Discussion). AN: Angola; MZ: Mozambique; EG: Equatorial Guinea; UG: Uganda; PT: Portugal. The other abbreviations are equal to the presented in Table S7.

	AN	MZ	EG	UG	AA_WCA	NS_WCA	NK_WCA	Fulani	Pygmy	Khoisan	Luo	NK_EA	AA_EA	NS_EA	SSA	PT	NWE	Brahui	Plt	EAsian	Nahua	NA	Cc
AN	*																						
MZ	-0.0179	*																					
EG	0.0042	0.0101	*																				
UG	-0.0123	-0.0111	0.0018	*																			
AA_WCA	-0.0022	-0.0012	-0.0045	-0.0037	*																		
NS_WCA	-0.0015	0.0065	0.0026	-0.0062	0.0030	*																	
NK_WCA	0.0008	-0.0032	0.0190	-0.0039	0.0101	-0.0033	*																
Fulani	0.0125	0.0207	-0.0068	0.0097	-0.0015	0.0099	0.0315	*															
Pygmy	0.0025	-0.0029	0.0034	0.0028	-0.0031	0.0216	0.0186	0.0070	*														
Khoisan	0.0089	0.0149	-0.0089	0.0091	-0.0054	0.0178	0.0343	-0.0137	-0.0029	*													
Luo	0.0347	0.0379	0.0133	0.0378	0.0154	0.0617	0.0709	0.0055	0.0079	-0.0125	*												
NK_EA	0.0586	0.0758	0.0209	0.0587	0.0340	0.0650	0.0955	0.0071	0.0417	-0.0021	-0.0013	*											
AA_EA	0.0047	0.0165	-0.0029	0.0046	0.0013	-0.0006	0.0217	-0.0029	0.0140	-0.0019	0.0246	0.0231	*										
NS_EA	0.0271	0.0414	0.0023	0.0264	0.0114	0.0245	0.0539	-0.0046	0.0222	-0.0078	0.0088	0.0001	0.0035	*									
SSA	0.0142	0.0178	0.0115	0.0238	0.0089	0.0500	0.0531	0.0088	-0.0002	-0.0081	-0.0145	0.0167	0.0217	0.0157	*								
PT	0.0488	0.0770	0.0235	0.0484	0.0400	0.0261	0.0728	0.0159	0.0663	0.0242	0.0591	0.0211	0.0138	0.0108	0.0658	*							
NWE	0.0836	0.1139	0.0525	0.0756	0.0720	0.0394	0.0943	0.0469	0.1077	0.0642	0.1126	0.0615	0.0363	0.0407	0.1178	-0.0009	*						
Brahui	0.0448	0.0716	0.0094	0.0417	0.0293	0.0180	0.0693	-0.0022	0.0574	0.0066	0.0444	-0.0054	0.0015	-0.0092	0.0561	-0.0320	-0.0228	*					
Plt	0.0475	0.0701	0.0174	0.0397	0.0336	0.0122	0.0585	0.0113	0.0628	0.0234	0.0678	0.0292	0.0089	0.0101	0.0735	-0.0136	-0.0103	-0.0393	*				
EAsian	0.0855	0.1105	0.0378	0.0738	0.0582	0.0477	0.0991	0.0270	0.0872	0.0376	0.0738	0.0257	0.0288	0.0197	0.0883	-0.0039	0.0031	-0.0358	-0.0120	*			
Nahua	0.1575	0.1670	0.1051	0.1524	0.1153	0.1908	0.1974	0.0907	0.1083	0.0730	0.0320	0.0333	0.1114	0.0744	0.0587	0.1311	0.1940	0.1448	0.1691	0.1383	*		
NA	0.2380	0.2398	0.1807	0.2248	0.1862	0.2828	0.2687	0.1732	0.1725	0.1619	0.1081	0.1279	0.1861	0.1513	0.1219	0.2279	0.2962	0.3179	0.2903	0.2406	0.0104	*	
Cc	0.0573	0.0912	0.0422	0.0602	0.0592	0.0295	0.0803	0.0372	0.0911	0.0506	0.0939	0.0510	0.0274	0.0327	0.0963	-0.0039	-0.0055	-0.0208	-0.0086	0.0060	0.1716	0.2664	*

The significant values of genetic distances after Bonferroni's correction are labelled in red.

Annex 4. *Umami* taste supplementary material

Table S10 – Data and information collected for *TAS1R1* gene.

Paper	Publication	Population	Abbreviation	n	CG	CA	TG
Kim *et al.* 2006	Chem Senses	Cameroonian	Ca	20	0.75	0.2	0.05
Kim *et al.* 2006	Chem Senses	Chinese	Ch	10	1	0	0
Kim *et al.* 2006	Chem Senses	Hungarian	Hu	10	1	0	0
Kim *et al.* 2006	Chem Senses	Japanese	Ja	10	1	0	0
Kim *et al.* 2006	Chem Senses	Native American	NA	10	1	0	0
Kim *et al.* 2006	Chem Senses	Northern European	NE	10	1	0	0
Kim *et al.* 2006	Chem Senses	Pakistani	Pa	8	1	0	0
Kim *et al.* 2006	Chem Senses	Russian	Ru	10	1	0	0

Table S11 – Data and information collected for *TAS1R3* gene.

Paper	Publication	Population	Abbreviation	n	GGC	GGT	GAC	AGC
Kim *et al.* 2006	Chem Senses	Cameroonian	Ca	20	0.725	0.25	0	0.025
Kim *et al*. 2006	Chem Senses	Chinese	Ch	10	1	0	0	0
Kim *et al.* 2006	Chem Senses	Hungarian	Hu	10	1	0	0	0
Kim *et al.* 2006	Chem Senses	Japanese	Ja	10	0.9	0	0.1	0
Kim *et al.* 2006	Chem Senses	Native American	NA	10	1	0	0	0
Kim *et al.* 2006	Chem Senses	Northern European	NE	10	1	0	0	0
Kim *et al.* 2006	Chem Senses	Pakistani	Pa	8	1	0	0	0
Kim *et al.* 2006	Chem Senses	Russian	Ru	10	1	0	0	0

Table S12 – AMOVA groups regarding the continent of origin for *TAS1R1 and TAS1R3*.

Africa	Europe	Asia	America
Mozambique	European	Chinese	Native American
Uganda	Hungarian	Pakistani	
Cameroonian	Portugal	Japanese	
Equatorial Guinea	Russian		

Annex 5. Sweet taste supplementary material

Table S13 – Data and information collected for *TAS1R3* gene concerning sweet taste.

Paper	Publication	Population	n	CC	CT	TT
Fushan *et al.* 2009	Curr Biol	African	15	0.4	0.23	0.37
Fushan *et al.* 2009	Curr Biol	European	92	0.91	0.005	0.085
Fushan *et al.* 2009	Curr Biol	Asian	37	0.78	0	0.22

Table S14 – AMOVA groups regarding the continent of origin.

Africa	Europe	Asia
Mozambique	European	Asian
Angola	Portugal	
Equatorial Guinea		
African		